IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

Exact Random Coding Error Exponents for the Two-User Interference Channels

Wasim Huleihel and Neri Merhav

CCIT Report #880
March 2015

Electronics
Computers
Communications

DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL

CCIT Report #880 March 2015

Exact Random Coding Error Exponents for the Two-User Interference Channel

Wasim Huleihel and Neri Merhav
Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 3200003, ISRAEL
E-mail: {wh@tx, merhav@ee}.technion.ac.il

Abstract

This paper is about exact error exponents for the two-user interference channel under the random coding regime. Specifically, we first analyze the standard random coding ensemble, where the codebooks are comprised of independently and identically distributed (i.i.d.) codewords. For this ensemble, we focus on optimum decoding, which is in contrast to other, heuristic decoding rules that have been used in the literature (e.g., joint typicality decoding, treating interference as noise, etc.). The fact that the interfering signal is a codeword, and not an i.i.d. noise process, complicates the application of conventional techniques of performance analysis of the optimum decoder. Also, unfortunately, these conventional techniques result in loose bounds. Using analytical tools rooted in statistical physics, as well as advanced union bounds, we derive exact single-letter formulas for the random coding error exponents. We compare our results with the best known lower bound on the error exponent, and show that our exponents can be strictly better. It turns out that the methods employed in this paper, can also be used to analyze more complicated coding ensembles. Accordingly, as an example, using the same techniques, we find exact formulas for the error exponent associated with the Han-Kobayashi (HK) random coding ensemble, which is based on superposition coding.

Index Terms

Random coding, error exponent, interference channels, superposition coding, Han-Kobayashi scheme, statistical physics, optimal decoding, multiuser communication.

*This research was partially supported by The Israeli Science Foundation (ISF), grant no. 412/12.

I. INTRODUCTION

A. Previous Work

The two-user interference channel (IFC) models a general scenario of communication between two transmitters and two receivers (with no cooperation at either side), where each receiver decodes its intended message from an observed signal, which is interfered by the other user, and corrupted by channel noise. The information-theoretic analysis of this model has begun over more than four decades ago and has recently witnessed a resurgence of interest. Most of the previous work on multiuser communication, and specifically, on the IFC, has focused on obtaining inner and outer bounds to the capacity region (see, for example, [1, Ch. II.7]). In a nutshell, the study of this kind of channel was started in [2], and continued in [3], where simple inner and outer bounds to the capacity region were given. Then, in [4], by using the well-known superposition coding technique, the inner bound of [3] was strictly improved. In [5], various inner and outer bounds were obtained by transforming the IFC model into some multiple-access or broadcast channel. Unfortunately, the capacity region for the general interference channel is still unknown, although it has been solved for some very special cases [6, 7]. The best known inner bound is the Han-Kobayashi (HK) region, established in [8], and which will also be considered in this paper. Recently, it was shown [9] that the capacity region can be strictly larger than the HK region.

To our knowledge, [10, 11] are the only previous works which treat the error exponents for the IFC under optimal decoding. Specifically, [10] derives lower bounds on error exponents of random codebooks comprised of i.i.d. codewords uniformly distributed over a given type class, under maximum likelihood (ML) decoding at each user, that is, optimal decoding. Contrary to the error exponent analysis of other multiuser communication systems, such as the multiple access channel [12], the difficulty in analyzing the error probability of the optimal decoder for the IFC is due to statistical dependencies induced by the interfering signal. Indeed, for the IFC, the marginal channel determining each receiver's ML decoding rule is induced also by the codebook of the interfering user. This indeed extremely complicates the analysis, mostly because the interfering signal is a codeword and not an i.i.d. process. Another important observation, which was noticed in [10], is that the usual bounding techniques (e.g., Gallager's bounding technique) on the error probability fail to give tight results. To alleviate this problem, the authors of [10], combined some of the ideas from Gallager's bounding technique [13] to get an upper bound on the average probability of decoding error under ML decoding, the method of types [14], and used the method of distance enumerators, in the spirit of [15], which allows to avoid the use of Jensen's inequality in some steps.

B. Contributions

The main purpose of this paper is to extend the study of achievability schemes to the more refined analysis of error exponents achieved by the two users, similarly as in [10]. Specifically, we derive exact single-letter expressions for error exponents associated with the average error probability, for the finite-alphabet two-user IFC, under several random coding ensembles. The main contributions of this paper are as follows:

- Similarly as in recent works (see, e.g., [12, 16-19] and references therein) on the analysis of error exponents, we derive exact single-letter formulas for the random coding error exponents, and not merely bounds as in [10]. For the standard random coding ensemble, considered in Subsection III-B, we analyze the optimal decoder for each receiver, which is interested solely in its intended message. This is in contrast to usual decoding techniques analyzed for the IFC, in which each receiver decodes, in addition to its intended message, also part of (or all) the interfering codeword (that is, the other user's message), or other conventional achievability arguments [1, Ch. II.7], which are based on joint-typicality decoding, with restrictions on the decoder (such as, "treat interference as noise" or to "decode the interference"). This enables us to understand whether there is any significant degradation in performance due to the sub-optimality of the decoder. Also, since [10] also analyzed the optimal decoder, we compared our exact formulas with their lower bound, and show that our error exponent can be strictly better, which implies that the bounding technique in [10] is not tight.
- As was mentioned earlier, in [10] only random codebooks comprised of i.i.d. codewords (uniformly distributed over a type class) were considered. These ensembles are much simpler than the superposition codebooks of [8]. Unfortunately, it very tedious to analyze superposition codebooks using the methods of [10], and even if we do so, the tightness is questionable. In this paper, however, the new tools that we have derived enable us to: first, as was mentioned before, obtain the exact error exponents, and secondly, to analyze more involved random coding ensembles. Indeed, in Subsection III-C, we consider the coding ensemble used in HK achievability scheme [8], and derive the respective error exponents. We also discuss an ensemble of hierarchical/tree codes [20]. Finally, it is worthwhile to mention that the analytical formulas of our error exponents are less tedious than the lower bound of [10].
- The exact analysis of the error exponents, carried out in this paper, turns out to be much more difficult than in previous works on point-to-point and multiuser communication problems, see, e.g., [12, 16-19]. Specifically, we encounter two main difficulties in our analysis: First, typically, when analyzing the probability of error, the first step is to apply the union bound. Usually, for point-to-point systems, under

the random coding regime, the average error probability can be written as a union of pairwise independent error events. Accordingly, in this case, it is well-known that the truncated union bound is exponentially tight [21, Lemma A.2]. This is no longer the case, however, when considering multiuser systems, and in particular, the IFC. For the IFC, the events comprising the union are strongly dependent, especially due to the fact that we are considering the optimal decoder. Indeed, recall that the optimal decoder for the first user, for example, declares that a certain message was transmitted if this message maximizes the likelihood pertaining to the marginal channel. This marginal channel[1] is the average of the actual channel over the messages of the interfering user, and thus depends on the whole codebook of the that user. Accordingly, the overall error event is the union of an exponential number of error events where each event depends on the marginal channel, and thus on the codebook of the interfering user. To alleviate this difficulty, following the ideas of [12], we derived new exponentially tight upper and lower bounds on the probability of a union of events, which takes into account the dependencies among the events. The second difficulty that we have encountered in our analysis is that in contrast to previous works, applying the distance enumerator method [15] is not simple, due to the reason mentioned above. Using some methods from large deviations theory, we were able to tackle this difficulty.

- We believe that by using the techniques and tools derived in this paper, other multiuser systems, such as the IFC with mismatched decoding, the MAC [12], the broadcast channel, the relay channel, etc., and accordingly, other coding schemes, such as binning [16], and hierarchical codes [20], can be analyzed.

The paper is organized as follows. In Section II, we establish notation conventions. In Section III, we formalize the problem and assert the main theorems. Specifically, in Subsections III-B and III-C, we give the resulting error exponents under the standard random coding ensemble and the HK coding ensemble, respectively. Finally, Section IV is devoted to the proofs of our main results.

II. Notation Conventions

Throughout this paper, scalar random variables (RVs) will be denoted by capital letters, their sample values will be denoted by the respective lower case letters, and their alphabets will be denoted by the respective calligraphic letters, e.g. X, x, and $\mathcal{X}$, respectively. A similar convention will apply to random vectors of dimension n and their sample values, which will be denoted with the same symbols in the boldface font. We also use the notation X_i^j $(j > i)$ to designate the sequence of RVs $(X_i, X_{i+1}, \ldots, X_j)$. The set of all n-vectors with components taking values in a certain finite alphabet, will be denoted as

[1] The precise definition will be given in the sequel.

the same alphabet superscripted by n, e.g., $\mathcal{X}^n$. Generic channels will be usually denoted by the letters P, Q, or W. We shall mainly consider joint distributions of two RVs (X,Y) over the Cartesian product of two finite alphabets $\mathcal{X}$ and $\mathcal{Y}$. For brevity, we will denote any joint distribution, e.g. Q_{XY}, simply by Q, the marginals will be denoted by Q_X and Q_Y, and the conditional distributions will be denoted by $Q_{X|Y}$ and $Q_{Y|X}$. The joint distribution induced by Q_X and $Q_{Y|X}$ will be denoted by $Q_X \times Q_{Y|X}$, and a similar notation will be used when the roles of X and Y are switched.

The expectation operator will be denoted by $\mathbb{E}\left\{\cdot\right\}$, and when we wish to make the dependence on the underlying distribution Q clear, we denote it by $\mathbb{E}_Q\left\{\cdot\right\}$. Information measures induced by the generic joint distribution Q_{XY}, will be subscripted by Q, for example, $I_Q(X;Y)$ will denote the corresponding mutual information, etc. The divergence (or, Kullback-Liebler distance) between two probability measures Q and P will be denoted by $D(Q||P)$. The weighted divergence between two channels, $Q_{Y|X}$ and $P_{Y|X}$, with weight P_X, is defined as

$$D(Q_{Y|X}||P_{Y|X}|P_X) \triangleq \sum_{x\in\mathcal{X}} P_X(x) \sum_{y\in\mathcal{Y}} Q_{Y|X}(y|x) \log \frac{Q_{Y|X}(y|x)}{P_{Y|X}(y|x)}. \tag{1}$$

For a given vector $\boldsymbol{x}$, let $\hat{Q}_{\boldsymbol{x}}$ denote the empirical distribution, that is, the vector $\{\hat{Q}_{\boldsymbol{x}}(x),\ x\in\mathcal{X}\}$, where $\hat{Q}_{\boldsymbol{x}}(x)$ is the relative frequency of the letter x in the vector $\boldsymbol{x}$. Let $T(P_X)$ denote the type class associated with P_X, that is, the set of all sequences $\boldsymbol{x}$ for which $\hat{Q}_{\boldsymbol{x}} = P_X$. Similarly, for a pair of vectors $(\boldsymbol{x},\boldsymbol{y})$, the empirical joint distribution will be denoted by $\hat{Q}_{\boldsymbol{xy}}$, or simply by $\hat{Q}$, for short. All previously defined notation rules for regular distributions will also be used for empirical distributions.

The cardinality of a finite set $\mathcal{A}$ will be denoted by $|\mathcal{A}|$, its complement will be denoted by $\mathcal{A}^c$. The probability of an event $\mathcal{E}$ will be denoted by $\Pr\left\{\mathcal{E}\right\}$. The indicator function of an event $\mathcal{E}$ will be denoted by $\mathcal{I}\left\{\mathcal{E}\right\}$. For two sequences of positive numbers, $\{a_n\}$ and $\{b_n\}$, the notation $a_n \doteq b_n$ means that $\{a_n\}$ and $\{b_n\}$ are of the same exponential order, i.e., $n^{-1}\log a_n/b_n \to 0$ as $n\to\infty$, where logarithms are defined with respect to (w.r.t.) the natural basis, that is, $\log(\cdot) = \ln(\cdot)$. Finally, for a real number x, we denote $[x]_+ \triangleq \max\{0,x\}$.

III. Problem Formulation and Main Results

In this section, we present the model, the main results, and discuss them. We split this section into two subsections, where in each one, we consider a different coding ensemble. We start with a simple random coding ensemble where random codebooks comprised of i.i.d. codewords uniformly distributed over a type class. It is well-known [11] that this coding scheme can be improved by using superposition coding and introducing the notion of "private" and "common" messages (to be defined in the sequel).

Accordingly, in the second subsection, we consider the HK coding scheme [8], and derive the exact error exponents. Finally, we discuss other ensembles that can be analyzed using the same methods.

A. The IFC Model

Consider a two-user interference channel of two senders, two receivers, and a discrete memoryless channel (DMC), defined by a set of single-letter transition probabilities, $W_{Y_1Y_2|X_1X_2}(y_1y_2|x_1x_2)$, with finite input alphabets $\mathcal{X}_1, \mathcal{X}_2$ and finite output alphabets $\mathcal{Y}_1, \mathcal{Y}_2$. Here, each sender, $k \in \{1,2\}$, wishes to communicate an independent message m_k at rate R_k, and each receiver, $l \in \{1,2\}$, wishes to decode its respective message. Specifically, a $(M_1 \triangleq e^{nR_1}, M_2 \triangleq e^{nR_2}, n)$ code $\mathcal{C}_n$ consists of:

- Two message sets $\mathcal{M}_1 \triangleq \{0,\ldots,M_1-1\}$ and $\mathcal{M}_2 \triangleq \{0,\ldots,M_2-1\}$ for the first and second users, respectively.
- Two encoders, where for each $k \in \{1,2\}$, the k-th encoder assigns a codeword $\boldsymbol{x}_{k,i}$ to each message $i \in \mathcal{M}_k$.
- Two decoders, where each decoder $l \in \{1,2\}$ assigns an estimate $\hat{m}_l$ to m_l.

We assume that the message pair (m_1, m_2) is uniformly distributed over $\mathcal{M}_1 \times \mathcal{M}_2$. It is clear that the *optimal decoder* of the first user, for this problem, is given by

$$\hat{m}_1 = \arg\max_{i\in\mathcal{M}_1} P\left(\boldsymbol{y}_1|\boldsymbol{x}_{1,i}\right) \tag{2}$$

$$= \arg\max_{i\in\mathcal{M}_1} \frac{1}{M_2}\sum_{j=1}^{M_2-1} P\left(\boldsymbol{y}_1|\boldsymbol{x}_{1,i},\boldsymbol{x}_{2,j}\right) \tag{3}$$

where $P(\boldsymbol{y}_1|\boldsymbol{x}_{1,i},\boldsymbol{x}_{2,j})$ is the marginal channel defined as

$$P(\boldsymbol{y}_1|\boldsymbol{x}_{1,i},\boldsymbol{x}_{2,j}) \triangleq \prod_{k=1}^{n} W_{Y_1|X_1X_2}(y_{1k}|x_{1ik}x_{2jk}), \tag{4}$$

and

$$W_{Y_1|X_1X_2}(y_{1k}|x_{1ik}x_{2jk}) \triangleq \sum_{y_{2k}\in\mathcal{Y}_2} W_{Y_1Y_2|X_1X_2}(y_{1k}y_{2k}|x_{1ik}x_{2jk}). \tag{5}$$

The optimal decoder of the second user is defined similarly. Accordingly, the probability of error for the code $\mathcal{C}_n$ and for the first user, is defined as

$$P_{e,1}\left(\mathcal{C}_n\right) \triangleq \Pr\left\{\hat{m}_1 \neq m_1\right\}, \tag{6}$$

and similarly for the second user.

B. The Ordinary Random Coding Ensemble

In this subsection, we consider the ordinary random coding ensemble: For each $k \in \{1,2\}$, we select independently M_k codewords $\boldsymbol{x}_{k,i}$, for $i \in \mathcal{M}_k$, under the uniform distribution across the type class $T(P_{X_k})$, for a given distribution P_{X_k} on $\mathcal{X}_k$. Our goal is to assess the exact exponential rate of $\bar{P}_{e,1} \triangleq \mathbb{E}\{P_{e,1}(\mathcal{C}_n)\}$, where the average is over the code ensemble, that is,

$$E_1^*(R_1,R_2) \triangleq \liminf_{n\to\infty}\left[-\frac{1}{n}\log \bar{P}_{e,1}\right], \tag{7}$$

and similarly for the second user. Before stating the main result, which is a single-letter formula of $E^*(R_1,R_2)$, we define some quantities. Given a joint distribution $Q_{X_1X_2Y_1}$ over $\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{Y}_1$, we define:

$$f(Q_{X_1X_2Y_1}) \triangleq \mathbb{E}_Q\left[\log W_{Y_1|X_1X_2}(Y_1|X_1X_2)\right] \tag{8}$$

$$= \sum_{(x_1,x_2,y_1)\in\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{Y}_1} Q_{X_1X_2Y_1}(x_1,x_2,y)\log W_{Y_1|X_1X_2}(y_1|x_1x_2), \tag{9}$$

$$t_0(Q_{X_1Y_1}) \triangleq R_2 + \max_{\hat{Q}:\, \hat{Q}_{X_1Y_1}=Q_{X_1Y_1},\, I_{\hat{Q}}(X_2;X_1,Y_1)\leq R_2}\left[f(\hat{Q}) - I_{\hat{Q}}(X_2;X_1,Y_1)\right], \tag{10}$$

and

$$E_1(\tilde{Q}_{X_1X_2Y_1},Q_{X_1X_2Y_1}) \triangleq \min_{\hat{Q}:\, \hat{Q}_{X_1Y_1}=\tilde{Q}_{X_1Y_1},\, \hat{Q}\in\mathcal{L}(\tilde{Q}_{X_1X_2Y_1},Q_{X_1X_2Y_1})}\left[I_{\hat{Q}}(X_2;X_1,Y_1)-R_2\right]_+ \tag{11}$$

where

$$\mathcal{L}(\tilde{Q}_{X_1X_2Y_1},Q_{X_1X_2Y_1}) \triangleq \left\{\hat{Q}:\ f(\tilde{Q}_{X_1X_2Y_1}) \leq \max\left[f(\hat{Q}),t_0(Q_{X_1X_2Y_1}),f(Q_{X_1X_2Y_1})\right],\right.$$
$$\left.\max\left[f(\hat{Q}),t_0(Q_{X_1X_2Y_1}),f(Q_{X_1X_2Y_1})\right]-f(\hat{Q}) \leq \left[R_2 - I_{\hat{Q}}(X_2;X_1,Y_1)\right]_+\right\}. \tag{12}$$

Finally, we define:

$$\hat{E}_1(Q_{X_1X_2Y_1},R_2) \triangleq \min_{\tilde{Q}:\, \tilde{Q}_{X_2Y_1}=Q_{X_2Y_1}}\left[I_{\tilde{Q}}(X_1;X_2,Y_1)+E_1(\tilde{Q}_{X_1X_2Y_1},Q_{X_1X_2Y_1})\right], \tag{13}$$

$$\hat{E}_2(Q_{X_1X_2Y_1},R_2) \triangleq \min_{\tilde{Q}:\, \tilde{Q}_{X_2Y_1}=Q_{X_2Y_1}} E_1(\tilde{Q}_{X_1X_2Y_1},Q_{X_1X_2Y_1}), \tag{14}$$

and

$$E^*(Q_{X_1X_2Y_1},R_1,R_2) \triangleq \max\left\{\left[\hat{E}_1(Q_{X_1X_2Y_1},R_2)-R_1\right]_+, \hat{E}_2(Q_{X_1X_2Y_1},R_2)\right\}. \tag{15}$$

Our main result is the following.

Theorem 1 Let R_1 and R_2 be given, and let $E^*(R_1, R_2)$ be defined as in (7). Consider the ensemble of fixed composition codes of types P_{X_1} and P_{X_2}, for the first and second users, respectively. For a discrete memoryless two-user IFC, we have:

$$E_1^*(R_1, R_2) = \min_{Q_{Y_1|X_1X_2}:\, Q_{X_1X_2} = P_{X_1}P_{X_2}} \left[D(Q_{Y_1|X_1X_2}||W_{Y_1|X_1X_2}|P_{X_1} \times P_{X_2}) + E^*(Q_{X_1X_2Y_1}, R_1, R_2)\right]. \tag{16}$$

Several remarks on Theorem 1 are in order.

• Due to symmetry, the error exponent for the second user, that is, $E_2^*(R_1, R_2)$ is simply obtained from Theorem 1 by swapping the roles of X_1, Y_1, and R_1, with X_2, Y_2, and R_2, respectively.

• An immediate byproduct of Theorem 1 is finding the set of rates (R_1, R_2) for which $E_1^*(R_1, R_2) > 0$, namely, for which the probability of error vanishes exponentially as $n \to \infty$. It is not difficult to show that this set is given by:

$$\mathcal{R}_{\text{ordinary},1} = \{R_1 < I(X_1; Y_1)\} \cup \{\{R_1 + R_2 < I(X_1, X_2; Y_1)\} \cap \{R_1 < I(X_1; Y_1|X_2)\}\} \tag{17}$$

evaluated with $P_{X_1X_2Y_1} = P_{X_1} \times P_{X_2} \times W_{Y_1|X_1X_2}$. Fig. 1 demonstrates a qualitative description of this region. The interpretation is as follows: The corner point $(I(X_1; Y_1|X_2), I(X_2; Y_1))$ is achieved by first decoding the interference (the second user), canceling it, and then decoding the first user. The sum-rate constraint can be achieved by joint decoding the two users (similarly to MAC), and thus, obviously, also by our optimal decoder. Finally, the region $R_1 < I(X_1; Y_1)$ and $R_2 \geq I(X_2; Y_1|X_1)$ means that we decode the first user while treating the interference as noise. Evidently, from the perspective of the first decoder, which is interested only in the message that is emitted from the first sender, the second sender can use any rate, and thus there is no bound on R_2 whenever $R_1 < I(X_1; Y_1)$. Note that this region was also obtained in [10], but from a lower bound on the error exponent. Accordingly, this means that according to [10], the achievable rate could be larger. Our results, however, show that one cannot do better when standard random coding is applied. Notice that $\mathcal{R}_{\text{ach},1}$ is well-known to be contained in the HK region [11, 22].

• *Existence of a single code:* our result holds true on the average, where the averaging is done over the random choice of codebooks. It can be shown (see, for example, [23, p. 2924]) that there exists deterministic sequence of fixed composition codebooks of increasing block length n for which the same asymptotic error performance can be achieved for *both* users simultaneously.

• *On the proof:* it is instructive to discuss (in some more detail than earlier) one of the main difficulties in proving Theorem 1, which is customary to multiuser systems, such as the IFC. Without loss of

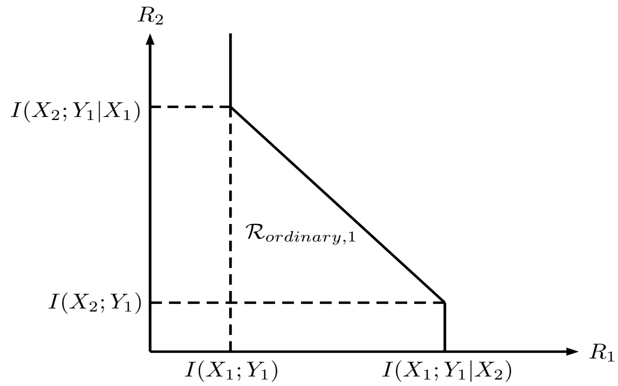

Fig. 1. Rate region $\mathcal{R}_{\text{ach},1}$ for which $E_1^*(R_1,R_2)>0$.

generality, we assume throughout, that the transmitted codewords are $\boldsymbol{x}_{1,0}$ and $\boldsymbol{x}_{2,0}$. Accordingly, the average probability of error associated with the decoder (3) is given by

$$\bar{P}_{e,1} = \Pr\left[\bigcup_{i=1}^{M_1-1}\left\{\sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}_1|\boldsymbol{X}_{1,i},\boldsymbol{X}_{2,j}\right) \geq \sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}_1|\boldsymbol{X}_{1,0},\boldsymbol{X}_{2,j}\right)\right\}\right] \tag{18}$$

$$= \mathbb{E}\left\{\Pr\left[\bigcup_{i=1}^{M_1-1}\left\{\sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}_1|\boldsymbol{X}_{1,i},\boldsymbol{X}_{2,j}\right) \geq \sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}_1|\boldsymbol{X}_{1,0},\boldsymbol{X}_{2,j}\right)\right\}\middle|\mathcal{F}_0\right]\right\} \tag{19}$$

where $\mathcal{F}_0 \triangleq (\boldsymbol{X}_{1,0},\boldsymbol{X}_{2,0},\boldsymbol{Y}_1)$. By the union bound and Shulman's inequality [21, Lemma A.2], we know that for a sequence of pairwise independent events, $\{\mathcal{A}_i\}_{i=1}^N$, the following holds

$$\frac{1}{2}\min\left\{1,\sum_{i=1}^N \Pr\left\{\mathcal{A}_i\right\}\right\} \leq \Pr\left\{\bigcup_{i=1}^N \mathcal{A}_i\right\} \leq \min\left\{1,\sum_{i=1}^N \Pr\left\{\mathcal{A}_i\right\}\right\}, \tag{20}$$

which is a useful result when assessing the exponential behavior of such probabilities. Equation (20) is one of the building blocks of tight exponential analysis of previously considered point-to-point systems (see, e.g., [16-19], and many references therein). However, it is evident that in our case the various events are not pairwise independent, and therefore this result cannot be applied directly. Indeed, since we are interested in the optimal decoder, each event of the union in (19), depends on the whole codebook of the second user. One may speculate that this problem can be tackled by conditioning on the codebook of the second user, and then (20). However, the cost of this conditioning is a very complicated (if not intractable) large deviations analysis of some quantities. To alleviate this problem, we derived new exponentially tight upper and lower bounds on the probability of union of events, which takes into account the dependencies among the events. This was done using the techniques of [12].

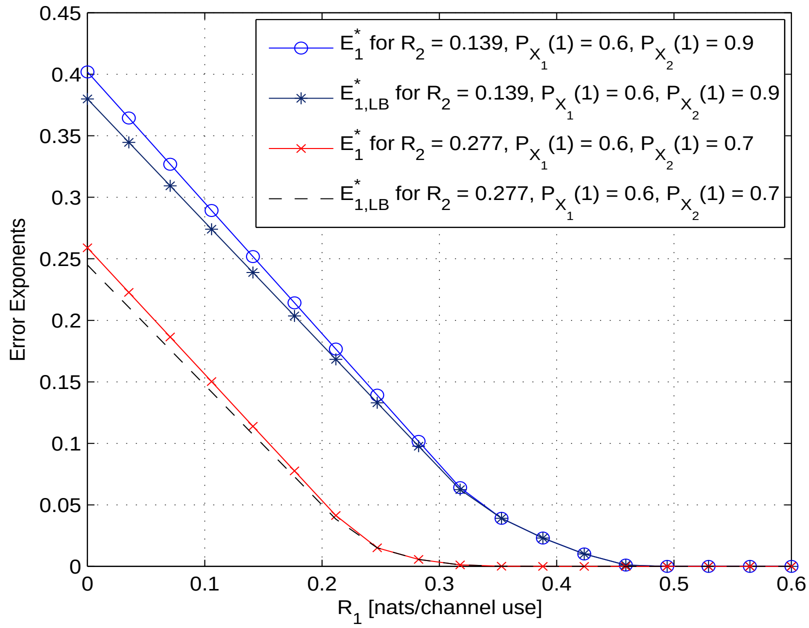

Fig. 2. Comparison between our error exponent $E_1^*(R_1,R_2)$ and the lower bound $E_{\text{LB}}(R_1,R_2)$ of [10], as a function of R_1 for two different values of R_2 and fixed choices of P_{X_1} and P_{X_2}.

- *Comparison with [10]*: Similarly to [10], we present results for the binary Z-channel model defined as follows: $Y_1 = X_1 \cdot X_2 \oplus Z$ and $Y_2 = X_2$, where $X_1, X_2, Y_1, Y_2 \in \{0,1\}$, $Z \sim \text{Bern}(p)$, "$\cdot$" is multiplication, and "$\oplus$" is modulo-2 addition. In the numerical calculations, we fix $p = 0.01$. Fig. 2 presents the exact error exponents under optimal decoding, derived in this paper, compared to the lower bound $E_{\text{LB}}(R_1,R_2)$ of [10], as a function of R_1, for different values of P_{X_1}, P_{X_2}, and R_2. It can be seen that our exponents can be strictly better than those of [10].

C. The Han-Kobayashi Coding Scheme

Consider the channel model of Subsection III-B. The best known inner bound on the capacity region is achieved by the HK coding scheme [8]. The idea of this scheme is to split the message M_1 into "private" and "common" messages, M_{11} and M_{12} at rates R_{11} and R_{12}, respectively, such that $R_1 = R_{11} + R_{12}$. Similarly M_2 is split into M_{21} and M_{22} at rates R_{21} and R_{22}, respectively, such that $R_2 = R_{21} + R_{22}$. The intuition behind this splitting is based on the receiver behavior at low and high signal-to-noise-ratio (SNR). Specifically, it is well-known [1] that: (1) when the SNR is low, treating the interference as noise is an optimal strategy, and (2) when the SNR is high, decoding and then canceling the interference is the optimal strategy. Accordingly, the above splitting captures the general intermediate situation where the first decoder, for example, is interested only in partial information from the second user, in addition to its own intended message.

Next, we describe explicitly the coding strategy, which was used in [8]. Fix a distribution $P_{Z_{11}}P_{Z_{12}}P_{Z_{21}}P_{Z_{22}}P_{X_1|Z_{11}Z_{12}}P_{X_2|Z_{21}Z_{22}}$, where the latter two conditional distributions represent deterministic mappings. For each $k,k' \in \{1,2\}$, randomly and conditionally independently generate a sequence $\boldsymbol{z}_{k,k'}(m_{k,k'})$ under the uniform distribution across the type class $T(P_{Z_{kk'}})$ for a given $P_{Z_{k,k'}}$. To communicate a message pair (m_{11},m_{12}), sender 1 transmits $\boldsymbol{x}_1(\boldsymbol{z}_{11},\boldsymbol{z}_{12})$, and analogously for sender 2. All our results can be extended to the setting in which the codewords are generated conditionally on a time-sharing sequence $\boldsymbol{q}$. However, this leads to more complex notation. Thus, we focus primarily on the case without time-sharing.

Let us now describe the operation of each receiver. Receiver $k=1,2$, recovers its intended message M_k and the common message from the other sender (although it is not required to). This scheme is illustrated in Fig. 3. Note that this decoding operation is the one that was used in [8], but there, the sub-optimal non-unique simultaneous joint typical decoder was used. Here, in contrast, we use sub-optimal ML decoding (the sub-optimality is due to the fact that our decoder recovers also the common message from the other sender). It is important to emphasize here that it was shown in [22] that optimal decoding, that is, the ML decoder that is interested only on its intended message, do not improves the achievable region. In other words, the HK achievable region cannot be improved upon merely by using optimal decoding. Nonetheless, in terms of error exponents, there could be an improvement.

We wish to find exact single-letter formulas for the error exponent, achieved by the HK encoding functions, in conjunction with the above described decoding functions. To this end, note that by combining the channel and the deterministic mappings as indicated by the dashed box in Fig. 3, the channel $(Z_{11},Z_{12},Z_{21},Z_{22}) \mapsto (Y_1,Y_2)$ is just a four-sender, two-receiver, DMC interference channel, with virtual inputs. We assume that the message quadruple $(M_{11},M_{12},M_{21},M_{22})$ is uniformly distributed over $\mathcal{M}_{11}\times\mathcal{M}_{12}\times\mathcal{M}_{21}\times\mathcal{M}_{22}$. Following the above descriptions, our decoder for this problem is given by

$$(\hat{m}_{11},\hat{m}_{12},\hat{m}_{21}) = \arg\max_{(i,j,k)\in\mathcal{M}_{11}\times\mathcal{M}_{12}\times\mathcal{M}_{21}} P\left(\boldsymbol{y}_1|\boldsymbol{z}_{11,i},\boldsymbol{z}_{12,j},\boldsymbol{z}_{21,k}\right) \tag{21}$$

$$= \arg\max_{(i,j,k)\in\mathcal{M}_{11}\times\mathcal{M}_{12}\times\mathcal{M}_{21}} \frac{1}{M_{22}}\sum_{l=0}^{M_{22}-1} P\left(\boldsymbol{y}_1|\boldsymbol{z}_{11,i},\boldsymbol{z}_{12,j},\boldsymbol{z}_{21,k},\boldsymbol{z}_{22,l}\right). \tag{22}$$

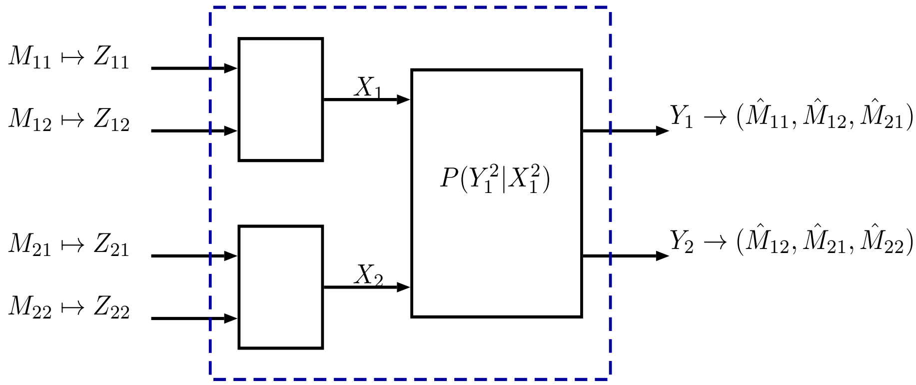

Fig. 3. Han-Kobayashi coding scheme.

Accordingly, the probability of error for the code $\mathcal{C}_n$ and for the first user, is defined as

$$P_{e,1}(\mathcal{C}_n) \triangleq \Pr\left\{(\hat{m}_{11}, \hat{m}_{12}, \hat{m}_{21}) \neq (m_{11}, m_{12}, m_{21})\right\}, \tag{23}$$

and similarly for the second user. Our goal is to assess the exact exponential rate of $\bar{P}_{e,1} \triangleq \mathbb{E}\left\{P_{e,1}(\mathcal{C}_n)\right\}$, where the average is over the code ensemble, namely,

$$E_{\text{HK}}^*(R_1, R_2) \triangleq \liminf_{n\to\infty}\left[-\frac{1}{n}\log \bar{P}_{e,1}\right], \tag{24}$$

and similarly for the second user.

We need some definitions. For simplicity of notation, in the following, we use the indexes $\{1,2,3,4\}$ instead of $\{11,12,21,22\}$, respectively. Let $\boldsymbol{Z} \triangleq (Z_1, Z_2, Z_3)$, and $\mathcal{U} \triangleq \{1,2,3,12,13,23,123\}$. For $u \in \{1,2,\ldots,7\}$, $\boldsymbol{Z}_{\mathcal{U}(u)}$ is a random vector consisting of the random variables which corresponds to the indexes in $\mathcal{U}(u)$, for example, $\boldsymbol{Z}_1 \triangleq \boldsymbol{Z}_{\mathcal{U}(1)} = Z_1$, $\boldsymbol{Z}_{12} \triangleq \boldsymbol{Z}_{\mathcal{U}(4)} = (Z_1, Z_2)$, $\boldsymbol{Z}_{123} \triangleq \boldsymbol{Z}_{\mathcal{U}(7)} = (Z_1, Z_2, Z_3)$, and so on. Define:

$$f(Q_{Z_1^4 Y_1}) \triangleq \mathbb{E}_Q\left[\log W_{Y_1|X_1(Z_1,Z_2)X_2(Z_3,Z_4)}(Y_1|X_1X_2)\right]. \tag{25}$$

Also, let

$$r_0(Q_{Z_1^3 Y_1}) \triangleq R_{22} + \max_{\hat{Q}:\, \hat{Q}_{Z_1^3 Y_1} = Q_{Z_1^3 Y_1},\, I_{\hat{Q}}(X_2; X_1, Y_1) \leq R_{22}} \left[f(\hat{Q}) - I_{\hat{Q}}(Z_4; Z_1^3, Y_1)\right], \tag{26}$$

and

$$E_1(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y_1}) \triangleq \min_{\hat{Q}:\, \hat{Q}_{Z_1^3 Y_1} = \tilde{Q}_{Z_1^3 Y_1},\, \hat{Q} \in \mathcal{D}(\tilde{Q}_{Z_1^4 Y_1}, Q_{Z_1^4 Y_1})} \left[I_{\hat{Q}}(Z_4; Z_1^3, Y_1) - R_{22}\right]_+, \tag{27}$$

where

$$\mathcal{D}(\tilde{Q}_{Z_1^4Y_1}, Q_{Z_1^4Y_1}) \triangleq \Big\{\hat{Q}: \ \max\Big[f(\hat{Q}), r_0(Q_{Z_1^4Y_1}), f(Q_{Z_1^4Y_1})\Big] - f(\hat{Q}) \le \Big[R_{22} - I_{\hat{Q}}(X_2; X_1, Y_1)\Big]_+, \\ f(\tilde{Q}_{Z_1^4Y_1}) \le \max\Big[f(\hat{Q}), r_0(Q_{Z_1^4Y_1}), f(Q_{Z_1^4Y_1})\Big]\Big\}. \tag{28}$$

Define:

$$\mathcal{R}_1 \triangleq R_{11};\ \mathcal{R}_2 \triangleq R_{12};\ \mathcal{R}_3 \triangleq R_{21};\ \mathcal{R}_4 \triangleq R_{11}+R_{12}; \\ \mathcal{R}_5 \triangleq R_{11}+R_{21};\ \mathcal{R}_6 \triangleq R_{12}+R_{21};\ \mathcal{R}_7 \triangleq R_{11}+R_{12}+R_{21}. \tag{29}$$

Now, let

$$\hat{E}^{(1)}(Q_{Z_1^4Y_1}, R_{22}) = \min_{\tilde{Q}:\ \tilde{Q}_{Z_2^4Y_1}=Q_{Z_2^4Y_1}} \Big[I_{\tilde{Q}}(Z_1; Z_2^4, Y_1) + E_1(\tilde{Q}_{Z_1^4Y_1}, Q_{Z_1^4Y_1})\Big], \tag{30}$$

$$\hat{E}^{(2)}(Q_{Z_1^4Y_1}, R_{22}) = \min_{\tilde{Q}:\ \tilde{Q}_{Z_1Z_3^4Y_1}=Q_{Z_1Z_3^4Y_1}} \Big[I_{\tilde{Q}}(Z_2; Z_1, Z_3^4, Y_1) + E_1(\tilde{Q}_{Z_1^4Y_1}, Q_{Z_1^4Y_1})\Big], \tag{31}$$

$$\hat{E}^{(3)}(Q_{Z_1^4Y_1}, R_{22}) = \min_{\tilde{Q}:\ \tilde{Q}_{Z_1^2Z_4Y_1}=Q_{Z_1^2Z_4Y_1}} \Big[I_{\tilde{Q}}(Z_3; Z_1^2, Z_4, Y_1) + E_1(\tilde{Q}_{Z_1^4Y_1}, Q_{Z_1^4Y_1})\Big], \tag{32}$$

and

$$\hat{E}_8^{(1)}(Q_{Z_1^4Y_1}, R_{22}) = \min_{\tilde{Q}:\ \tilde{Q}_{Z_2^4Y_1}=Q_{Z_2^4Y_1}} E_1(\tilde{Q}_{Z_1^4Y}, Q_{Z_1^4Y_1}), \tag{33}$$

$$\hat{E}_8^{(2)}(Q_{Z_1^4Y_1}, R_{22}) = \min_{\tilde{Q}:\ \tilde{Q}_{Z_1Z_3^4Y_1}=Q_{Z_1Z_3^4Y_1}} E_1(\tilde{Q}_{Z_1^4Y}, Q_{Z_1^4Y_1}), \tag{34}$$

$$\hat{E}_8^{(3)}(Q_{Z_1^4Y_1}, R_{22}) = \min_{\tilde{Q}:\ \tilde{Q}_{Z_1^2Z_4Y_1}=Q_{Z_1^2Z_4Y_1}} E_1(\tilde{Q}_{Z_1^4Y}, Q_{Z_1^4Y_1}). \tag{35}$$

For $u \in \{1,2,4\}$, let

$$\hat{E}_u^{(4)}(Q_{Z_1^4Y_1}, R_{22}) = \min_{\tilde{Q}:\ \tilde{Q}_{Z_3^4Y_1}=Q_{Z_3^4Y_1}} \Big[I_{\tilde{Q}}(\boldsymbol{Z}_{\mathcal{U}(u)}; Z_3^4, Y_1 | \boldsymbol{Z}_{12\setminus\mathcal{U}(u)}) + E_1(\tilde{Q}_{Z_1^4Y_1}, Q_{Z_1^4Y})\Big], \tag{36}$$

$$\hat{E}_8^{(4)}(Q_{Z_1^4Y_1}, R_{22}) = \min_{\tilde{Q}:\ \tilde{Q}_{Z_3^4Y_1}=Q_{Z_3^4Y_1}} E_1(\tilde{Q}_{Z_1^4Y}, Q_{Z_1^4Y_1}). \tag{37}$$

For $u \in \{1,3,5\}$:

$$\hat{E}_u^{(5)}(Q_{Z_1^4Y_1}, R_{22}) = \min_{\tilde{Q}:\ \tilde{Q}_{Z_2Z_4Y_1}=Q_{Z_2Z_4Y_1}} \Big[I_{\tilde{Q}}(\boldsymbol{Z}_{\mathcal{U}(u)}; Z_2, Z_4, Y_1 | \boldsymbol{Z}_{13\setminus\mathcal{U}(u)}) + E_1(\tilde{Q}_{Z_1^4Y_1}, Q_{Z_1^4Y_1})\Big], \tag{38}$$

$$\hat{E}_8^{(5)}(Q_{Z_1^4Y_1}, R_{22}) = \min_{\tilde{Q}:\ \tilde{Q}_{Z_2Z_4Y_1}=Q_{Z_2Z_4Y_1}} E_1(\tilde{Q}_{Z_1^4Y_1}, Q_{Z_1^4Y_1}). \tag{39}$$

For $u \in \{2,3,6\}$:

$$\hat{E}_u^{(5)}(Q_{Z_1^4Y_1}, R_{22}) = \min_{\tilde{Q}:\ \tilde{Q}_{Z_1Z_4Y_1}=Q_{Z_1Z_4Y_1}} \Big[I_{\tilde{Q}}(\boldsymbol{Z}_{\mathcal{U}(u)}; Z_1, Z_4, Y_1 | \boldsymbol{Z}_{13\setminus\mathcal{U}(u)}) + E_1(\tilde{Q}_{Z_1^4Y_1}, Q_{Z_1^4Y_1})\Big], \tag{40}$$

$$\hat{E}_8^{(6)}(Q_{Z_1^4Y_1}, R_{22}) = \min_{\tilde{Q}\colon \tilde{Q}_{Z_1Z_4Y_1}=Q_{Z_1Z_4Y_1}} E_1(\tilde{Q}_{Z_1^4Y_1}, Q_{Z_1^4Y_1}). \tag{41}$$

For $u \in \{1, 2, \ldots, 7\}$:

$$\hat{E}_u^{(7)}(Q_{Z_1^4Y_1}, R_{22}) = \min_{\tilde{Q}\colon \tilde{Q}_{Z_4Y_1}=Q_{Z_4Y_1}} \left[I_{\tilde{Q}}(\boldsymbol{Z}_{\mathcal{U}(u)}; Z_4, Y_1|\boldsymbol{Z}_{123\setminus\mathcal{U}(u)}) + E_1(\tilde{Q}_{Z_1^4Y_1}, Q_{Z_1^4Y_1})\right], \tag{42}$$

$$\hat{E}_8^{(7)}(Q_{Z_1^4Y_1}, R_{22}) = \min_{\tilde{Q}\colon \tilde{Q}_{Z_4Y_1}=Q_{Z_4Y_1}} E_1(\tilde{Q}_{Z_1^4Y_1}, Q_{Z_1^4Y_1}). \tag{43}$$

Finally, for $u \in \{1, 2, 3\}$, let

$$E_{\text{HK}}^{(u)}(Q_{Z_1^4Y_1}) \triangleq \max\left\{\left[\hat{E}^{(u)}(Q_{Z_1^4Y_1}, R_{22}) - \mathcal{R}_u\right]_+, \hat{E}_8^{(u)}(Q_{Z_1^4Y_1}, R_{22})\right\}, \tag{44}$$

$$E_{\text{HK}}^{(4)}(Q_{Z_1^4Y_1}) \triangleq \max\left\{\max_{u\in\{1,2,4\}}\left[\hat{E}_u^{(4)}(Q_{Z_1^4Y_1}, R_{22}) - \mathcal{R}_u\right]_+, \hat{E}_8^{(4)}(Q_{Z_1^4Y_1}, R_{22})\right\}, \tag{45}$$

$$E_{\text{HK}}^{(5)}(Q_{Z_1^4Y_1}) \triangleq \max\left\{\max_{u\in\{1,3,5\}}\left[\hat{E}_u^{(5)}(Q_{Z_1^4Y_1}, R_{22}) - \mathcal{R}_u\right]_+, \hat{E}_8^{(5)}(Q_{Z_1^4Y_1}, R_{22})\right\}, \tag{46}$$

$$E_{\text{HK}}^{(6)}(Q_{Z_1^4Y_1}) \triangleq \max\left\{\max_{u\in\{2,3,6\}}\left[\hat{E}_u^{(6)}(Q_{Z_1^4Y_1}, R_{22}) - \mathcal{R}_u\right]_+, \hat{E}_8^{(6)}(Q_{Z_1^4Y_1}, R_{22})\right\}, \tag{47}$$

$$E_{\text{HK}}^{(7)}(Q_{Z_1^4Y_1}) \triangleq \max\left\{\max_{u\in\{1:7\}}\left[\hat{E}_u^{(7)}(Q_{Z_1^4Y_1}, R_{22}) - \mathcal{R}_u\right]_+, \hat{E}_8^{(7)}(Q_{Z_1^4Y_1}, R_{22})\right\}. \tag{48}$$

Our second main result is the following.

Theorem 2 Let R_{11}, R_{12}, R_{21} and R_{22} be given such that $R_1 = R_{11} + R_{12}$ and $R_2 = R_{21} + R_{22}$, and let $E_{\text{HK}}^*(R_1, R_2)$ be defined as in (24). Consider the HK encoding scheme described above. For a discrete memoryless two-user IFC, we have:

$$E_{\text{HK}}^*(R_1, R_2) = \min_{Q_{Y_1|Z_1^4}\colon Q_{Z_1^4}=P_{Z_1^4}} \left[D(Q_{Y_1|Z_1^4}||W_{Y_1|Z_1^4}|P_{Z_1^4}) + \min_{u\in\{1:7\}} E_{\text{HK}}^{(u)}(Q_{Z_1^4Y_1})\right]. \tag{49}$$

Several remarks on Theorem 2 are in order.

- As before, an immediate byproduct of Theorem 2 is finding the set of rates (R_1, R_2) for which $E_1^*(R_1, R_2) > 0$, namely, for which the probability of error vanishes exponentially as $n \to \infty$. It can be shown that this set is given by the HK region, that is,

$$R_{11} \leq I(Z_1; Y_1|Z_2, Z_3), \tag{50a}$$

$$R_{12} \leq I(Z_2; Y_1|Z_1, Z_3), \tag{50b}$$

$$R_{21} \leq I(Z_3; Y_1|Z_1, Z_2), \tag{50c}$$

$$R_{11} + R_{12} \leq I(Z_1, Z_2; Y_1|Z_3), \tag{50d}$$

$$R_{11} + R_{21} \leq I(Z_1, Z_3; Y_1|Z_2), \tag{50e}$$

$$R_{12}+R_{21}\leq I(Z_2,Z_3;Y_1|Z_1), \tag{50f}$$

$$R_{11}+R_{12}+R_{21}\leq I(Z_1,Z_2,Z_3;Y_1), \tag{50g}$$

evaluated with $P_{Z_1^4Y_1}=P_{Z_1}P_{Z_2}P_{Z_3}P_{Z_4}W_{Y_1|X_1(Z_1,Z_2)X_2(Z_3,Z_4)}$, and similarly for the second user. As was mentioned earlier, it is possible to introduce a time-sharing sequence $\boldsymbol{q}$, and accordingly, (50) remains almost the same, but with some time-sharing random variable Q, appearing at the conditioning of each the above mutual information terms. Also, it can be shown that the above region includes $\mathcal{R}_{\text{ordinary},1}$, and thus, the HK ensemble is obviously better than the standard random coding ensemble described in Subsection III-B. Finally, it can be seen that using the ML decoder instead of the non-unique simultaneous joint typical decoder [8] cannot improve the achievable region (but will certainly improve the error exponent). This result is consistent with [22], where this fact was implied from another point of view.

• Using the same techniques and tools derived in this paper, we can consider other random coding ensembles. For example, we can analyze the error exponents resulting from the *hierarchical code ensemble*. Specifically, in this ensemble, the message M_1 is split into a common and private messages M_{11}, M_{12} at rates R_{11} and R_{12}, respectively, such that $R_1=R_{11}+R_{12}$. Similarly M_2 is split into a common and private messages M_{21}, M_{22} at rates R_{21} and R_{22}, respectively, such that $R_2=R_{21}+R_{22}$. Then, we first randomly draw a rate R_{11} codebook of block length n according to a given distribution. Then, for each such codeword, we randomly and conditionally independently generate a rate R_{12} codebook of block length n. In other words, the code has a tree structure with two levels, where the first serves for "cloud centers", and the second for the "satellites". We do the same for the second user. Under this ensemble, we can analyze the optimal decoder. Note, however, that this ensemble is different from the product ensemble considered in Theorem 2. Indeed, while for the former for each first stage codeword (cloud center) we independently draw a new codebook (satellites), for the latter, for each cloud center we have the same satellite. Loosely speaking, this means that the product ensemble is "less random". From the point of view of achievable region, however, the hierarchical ensemble is equivalent to the product ensemble used in HK scheme [1, Ch. II.7].

• In Theorem 2 we assumed the sub-optimal decoder given in (22). Indeed, the optimal decoder for our problem is given by:

$$(\hat{m}_{11},\hat{m}_{12})=\arg\max_{i,j}P\left(\boldsymbol{y}_1|\boldsymbol{z}_{11,i},\boldsymbol{z}_{12,j}\right) \tag{51}$$

$$=\arg\max_{i,j}\frac{1}{M_{21}M_{22}}\sum_{k=0}^{M_{21}-1}\sum_{l=0}^{M_{22}-1}P\left(\boldsymbol{y}_1|\boldsymbol{z}_{11,i},\boldsymbol{z}_{12,j},\boldsymbol{z}_{21,k},\boldsymbol{z}_{22,l}\right). \tag{52}$$

Unfortunately, it turns out that analyzing the HK scheme (in conjunction with (52)) is much more difficult, and requires some more delicate tools from large deviations theory. Specifically, the main difficulty in the derivations, is to analyze the large deviations behavior of a two-dimensional sum (due to the double summation in (52)) involving binomial random variables which are strongly dependent (contrary to the standard one-dimensional version, see, e.g., [16, p. 6027-6028]). Nonetheless, we note that for the hierarchical code ensemble described above, the optimal decoder can be analyzed. Indeed, for this ensemble, it is clear that the optimal decoder is given by

$$(\hat{m}_{11},\hat{m}_{12}) = \arg\max_{i,j} P\left(\boldsymbol{y}_1|\boldsymbol{x}_1(i,j)\right) \tag{53}$$

$$= \arg\max_{i,j} \frac{1}{M_{21}M_{22}}\sum_{k=0}^{M_{21}-1}\sum_{l=0}^{M_{22}-1} P\left(\boldsymbol{y}_1|\boldsymbol{x}_1(i,j),\boldsymbol{x}_2(k,l)\right) \tag{54}$$

where $\boldsymbol{x}_1(i,j) \triangleq f_1(\boldsymbol{x}_1'(i),\boldsymbol{x}_1''(i,j))$ and $\boldsymbol{x}_2(i,j) \triangleq f_2(\boldsymbol{x}_2'(i),\boldsymbol{x}_2''(i,j))$ due to the hierarchical structure. Now, while here too, we will deal with two-dimensional summation, the summands will be independent, given the cloud centers codebook, and the proof can be carried out smoothly.

IV. Proofs

A. *Proof of Theorem 1:*

Without loss of generality, we assume throughout, that the transmitted codewords are $\boldsymbol{x}_{1,0}$ and $\boldsymbol{x}_{2,0}$, and due to the fact that we analyze the first decoder, for convenience, we use $\boldsymbol{y}$ instead of $\boldsymbol{y}_1$. Accordingly, the average probability of error associated with the optimal decoder (3), is given by

$$P_e = \Pr\left[\bigcup_{i=1}^{M_1-1}\left\{\sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,i},\boldsymbol{X}_{2,j}\right) \geq \sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,0},\boldsymbol{X}_{2,j}\right)\right\}\right] \tag{55}$$

$$= \mathbb{E}\left\{\Pr\left[\bigcup_{i=1}^{M_1-1}\left\{\sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,i},\boldsymbol{X}_{2,j}\right) \geq \sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,0},\boldsymbol{X}_{2,j}\right)\right\}\middle|\mathcal{F}_0\right]\right\} \tag{56}$$

where $\mathcal{F}_0 \triangleq (\boldsymbol{X}_{1,0},\boldsymbol{X}_{2,0},\boldsymbol{Y})$. In the following, we propose new upper and lower bounds on the probability of a union of events, which are tight in the exponential scale, and suitable for some structured dependency between the events, as above. Before doing that, in order to give some motivation for these new bounds, we first rewrite (55) in another (equivalent) form. Specifically, we express (56) in terms of the joint types of $(\boldsymbol{X}_{1,0},\boldsymbol{X}_{2,0},\boldsymbol{Y})$ and $\{(\boldsymbol{Y},\boldsymbol{X}_{1,i},\boldsymbol{X}_{2,j})\}_{i,j}$. First, for a given joint distribution $Q_{X_1X_2Y}$ of $(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{y})$, we let

$$f\left(Q_{X_1X_2Y}\right) \triangleq \frac{1}{n}\log P\left(\boldsymbol{y}|\boldsymbol{x}_1,\boldsymbol{x}_2\right) \tag{57}$$

$$= \mathbb{E}_Q \left[\log W_{Y|X_1X_2}(Y|X_1X_2) \right]. \tag{58}$$

Now, for a given joint type $Q_{X_{1,0}X_{2,0}Y}$ of the random vectors $(\boldsymbol{X}_{1,0}, \boldsymbol{X}_{2,0}, \boldsymbol{Y})$, we define the set:

$$T_I\left(Q_{X_{1,0}X_{2,0}Y}\right) \triangleq \Bigg\{ \tilde{Q}^0_{X_1X_{2,0}Y} \in \mathcal{S}_0, \left(\left\{ \tilde{Q}^k_{X_1X_2Y} \right\}_{k=1}^{M_2-1}, \left\{ \hat{Q}^k_{X_{1,0}X_2Y} \right\}_{k=1}^{M_2-1} \right) \in \mathcal{S}_1 : \\ e^{nf(\tilde{Q}^0_{X_1X_{2,0}Y})} + \sum_{k=1}^{M_2-1} \left[e^{nf(\tilde{Q}^k_{X_1X_2Y})} - e^{nf(\hat{Q}^k_{X_1X_2Y})} \right] \geq e^{nf(Q_{X_{1,0}X_{2,0}Y})} \Bigg\} \tag{59}$$

where

$$\mathcal{S}_0(Q_{X_{1,0}X_{2,0}Y}) \triangleq \left\{ \tilde{Q}^0_{X_1X_{2,0}Y} : \tilde{Q}^0_{X_1} = P_{X_1}, \tilde{Q}^0_{X_2} = P_{X_2}, \tilde{Q}^0_{X_{2,0}Y} = Q_{X_{2,0}Y} \right\}, \tag{60}$$

and

$$\mathcal{S}_1(Q_{X_{1,0}X_{2,0}Y}) \triangleq \Bigg\{ \left\{ \tilde{Q}^k_{X_1X_2Y} \right\}_{k=1}^{M_2-1}, \left\{ \hat{Q}^k_{X_1X_2Y} \right\}_{k=1}^{M_2-1} : \tilde{Q}^k_{X_1} = P_{X_1}, \tilde{Q}^k_{X_2} = P_{X_2}, \tilde{Q}^k_Y = Q_Y, \\ \hat{Q}^k_{X_{1,0}} = P_{X_1}, \hat{Q}^k_{X_2} = P_{X_2}, \hat{Q}^k_{X_{1,0}Y} = Q_{X_{1,0}Y}, \ \forall 1 \leq k \leq M_2 - 1 \\ \tilde{Q}^k_{X_2Y} = \hat{Q}^k_{X_2Y}, \tilde{Q}^k_{X_1Y} = \tilde{Q}^m_{X_1Y}, \forall k, m \Bigg\}. \tag{61}$$

The set $T_I(Q_{X_{1,0}X_{2,0}Y})$ is the set of all possible types of $(\boldsymbol{X}_{1,i}, \mathcal{C}_2)$, where $\mathcal{C}_2$ denotes the codebook of the second user, which lead to a decoding error when $(\boldsymbol{X}_{1,0}, \boldsymbol{X}_{2,0}, \boldsymbol{Y}) \in T(Q_{X_{1,0}X_{2,0}Y})$ is transmitted. The various marginal constraints in (60) and (61) arise from the fact that we are assuming constant-composition random coding and, of course, fixed marginals due to the given fixed joint distribution $Q_{X_{1,0}X_{2,0}Y}$. Finally, the constraint:

$$e^{nf(\tilde{Q}^0_{X_1X_{2,0}Y})} + \sum_{k=1}^{M_2-1} \left[e^{nf(\tilde{Q}^k_{X_1X_2Y})} - e^{nf(\hat{Q}^k_{X_1X_2Y})} \right] \geq e^{nf(Q_{X_{1,0}X_{2,0}Y})} \tag{62}$$

in (59), represents a decoding error event, that is, it holds if and only if

$$\sum_{j=0}^{M_2-1} P\left(\boldsymbol{y}|\boldsymbol{x}_{1,i}, \boldsymbol{x}_{2,j}\right) \geq \sum_{j=0}^{M_2-1} P\left(\boldsymbol{y}|\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,j}\right) \tag{63}$$

for $(\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,0}, \boldsymbol{y}) \in T(Q_{X_{1,0}X_{2,0}Y})$, $(\boldsymbol{x}_{1,i}, \boldsymbol{x}_{2,0}, \boldsymbol{y}) \in T(\tilde{Q}^0_{X_1X_{2,0}Y})$, $\left\{ (\boldsymbol{x}_{1,i}, \boldsymbol{x}_{2,j}, \boldsymbol{y}) \in T(\tilde{Q}^j_{X_1X_2Y}) \right\}_{j=1}^{M_2-1}$, and $\left\{ (\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,j}, \boldsymbol{y}) \in T(\hat{Q}^j_{X_{1,0}X_2Y}) \right\}_{j=1}^{M_2-1}$. Now, with these definitions, fixing $Q_{X_{1,0}X_{2,0}Y}$, and letting $(\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,0}, \boldsymbol{y})$ be an arbitrary triplet of sequences such that $(\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,0}, \boldsymbol{y}) \in T(Q_{X_{1,0}X_{2,0}Y})$, it follows, by definition, that the error event

$$\bigcup_{i=1}^{M_1-1} \left\{ \sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,i}, \boldsymbol{X}_{2,j}\right) \geq \sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,0}, \boldsymbol{X}_{2,j}\right) \right\} \tag{64}$$

can be rewritten, in terms of types, as follows:

$$\bigcup_{i=1}^{M_1-1} \bigcup_{\left\{\tilde{Q}^j_{X_1X_2Y},\hat{Q}^j_{X_1X_2Y}\right\}_j \in T_I(Q_{X_{1,0}X_{2,0}Y})} \left\{ \begin{array}{c} (\boldsymbol{X}_{1,i},\boldsymbol{x}_{2,0},\boldsymbol{y}) \in T(\tilde{Q}^0_{X_1X_{2,0}Y}), \\ \left\{(\boldsymbol{X}_{1,i},\boldsymbol{X}_{2,j},\boldsymbol{y}) \in T(\tilde{Q}^j_{X_1X_2Y})\right\}_{j=1}^{M_2-1}, \\ \left\{(\boldsymbol{x}_{1,0},\boldsymbol{X}_{2,j},\boldsymbol{y}) \in T(\hat{Q}^j_{X_{1,0}X_2Y})\right\}_{j=1}^{M_2-1} \end{array} \right\}. \tag{65}$$

We wish to analyze the probability of the event in (65), conditioned on $\mathcal{F}_0$. Note that the inner union in (65) is over vectors of types (an exponential number of them). Finally, for the sake of convenience, we simplify the notations of (65), and write it equivalently as

$$\bigcup_{i=1}^{M_1-1} \bigcup_{\boldsymbol{l}} \left\{ \begin{array}{c} \boldsymbol{X}_{1,i} \in \mathcal{A}_{\boldsymbol{l},0}, \\ (\boldsymbol{X}_{1,i},\boldsymbol{X}_{2,j}) \in \mathcal{A}_{\boldsymbol{l},j}, \text{ for } j=1,\ldots,M_2-1, \\ \boldsymbol{X}_{2,j} \in \tilde{\mathcal{A}}_{\boldsymbol{l},j}, \quad \text{for } j=1,\ldots,M_2-1 \end{array} \right\} \tag{66}$$

where, again, the index "$\boldsymbol{l}$" in the inner union runs over the combinations of types (namely, $\boldsymbol{l} = \left\{\tilde{Q}^j_{X_1X_2Y},\hat{Q}^j_{X_1X_2Y}\right\}_j$) that belong to $T_I(Q_{X_{1,0}X_{2,0}Y})$, and the various sets $\left\{\mathcal{A}_{\boldsymbol{l},j},\tilde{\mathcal{A}}_{\boldsymbol{l},j}\right\}_{\boldsymbol{l},j}$ correspond to the typical sets in (65) (recall that $(\boldsymbol{x}_{1,0},\boldsymbol{x}_{2,0},\boldsymbol{y})$ are given at this stage). Next, following the ideas of [12], we provide exponentially tight lower and upper bounds on a generic probability which has the form of (66). The proof of this Lemma is relegated to Appendix A.

Lemma 1 Let $\{V_1(i)\}_{i=1}^{L_1}, V_2, V_3, \ldots, V_K$ be independent sequences of independently and identically distributed (i.i.d.) random variables on the alphabets $\mathcal{V}_1 \times \mathcal{V}_2 \times \ldots \times \mathcal{V}_K$, respectively, with $V_1(i) \sim P_{V_1}, V_2 \sim P_{V_2}, \ldots, V_K \sim P_{V_K}$. Fix a sequence of sets $\{\mathcal{A}_{i,1}\}_{i=1}^N, \{\mathcal{A}_{i,2}\}_{i=1}^N, \ldots, \{\mathcal{A}_{i,K-1}\}_{i=1}^N$, where $\mathcal{A}_{i,j} \subseteq \mathcal{V}_1 \times \mathcal{V}_{j+1}$, for $1 \leq j \leq K-1$ and for all $1 \leq i \leq N$. Also, fix a set $\{\mathcal{A}_{i,0}\}_{i=1}^N$ where $\mathcal{A}_{i,0} \subseteq \mathcal{V}_1$ for all $1 \leq i \leq N$, and another sequence of sets $\{\mathcal{G}_{i,2}\}_{i=1}^N, \{\mathcal{G}_{i,3}\}_{i=1}^N, \ldots, \{\mathcal{G}_{i,K}\}_{i=1}^N$, where $\mathcal{G}_{i,j} \subseteq \mathcal{V}_j$, for $2 \leq j \leq K$ and for all $1 \leq i \leq N$. Define

$$\mathcal{B}_{m,1} \triangleq \left\{ v_1 : \; v_1 \in \mathcal{A}_{l,0}, \; \bigcap_{j=1}^{K-1} (v_1, v_{j+1}) \in \mathcal{A}_{l,j}, \; \bigcap_{j=2}^{K} v_j \in \mathcal{G}_{l,j} \; \text{ for some } \; \{v_j\}_{j=2}^K \right\}, \tag{67}$$

and

$$\mathcal{B}_{m,2} \triangleq \left\{ \{v_j\}_{j=2}^K : \; v_1 \in \mathcal{A}_{l,0}, \; \bigcap_{j=1}^{K-1} (v_1, v_{j+1}) \in \mathcal{A}_{l,j}, \; \bigcap_{j=2}^{K} v_j \in \mathcal{G}_{l,j} \; \text{ for some } v_1 \right\}, \tag{68}$$

for $m = 1, 2, \ldots, N$. Then,

1) A general upper bound is given by

$$\Pr\left\{ \bigcup_i \left\{ \bigcup_{m=1}^N \left\{ V_1(i) \in \mathcal{A}_{l,0}, \; \bigcap_{k=1}^{K-1} (V_1(i), V_{k+1}) \in \mathcal{A}_{m,k}, \; \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{m,k} \right\} \right\} \right\}$$

$$\leq \min\left\{1, L_1 \Pr\left\{\bigcup_{m=1}^{N}\{V_1 \in \mathcal{B}_{m,1}\}\right\}, \Pr\left\{\bigcup_{m=1}^{N}\left\{\{V_j\}_{k=2}^{K} \in \mathcal{B}_{m,2}\right\}\right\},\right.$$

$$\left. L_1 \Pr\left\{\bigcup_{m=1}^{N}\left\{V_1 \in \mathcal{A}_{m,0}, \bigcap_{k=1}^{K-1}(V_1, V_{k+1}) \in \mathcal{A}_{m,k}, \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{m,k}\right\}\right\}\right\} \quad (69)$$

with $(V_1, \ldots, V_K) \sim P_{V_1} \cdots \times P_{V_K}$.

2) If $\{V_1(i)\}_{i=1}^{L_1}, V_2, V_3, \ldots, V_K$ are all independent, $\{V_1(i)\}_{i=1}^{L_1}$ is a sequence of pairwise independent and identically distributed random variables, and

$$\Pr\left\{\bigcup_{m=1}^{N}\left\{v_1 \in \mathcal{A}_{l,0}, \bigcap_{k=1}^{K-1}(v_1, V_{k+1}) \in \mathcal{A}_{m,k}, \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{m,k}\right\}\right\} \quad (70)$$

is the same for all $v_1 \in \mathcal{B}_{1,1}$, and for all $v_1 \in \mathcal{B}_{2,1}$, and so on till $v_1 \in \mathcal{B}_{N,1}$, but may be different for different $\mathcal{B}_{l,1}$, and

$$\Pr\left\{\bigcup_{m=1}^{N}\left\{V_1 \in \mathcal{A}_{l,0}, \bigcap_{k=1}^{K-1}(V_1, v_{k+1}) \in \mathcal{A}_{m,k}, \bigcap_{k=2}^{K} v_k \in \mathcal{G}_{m,k}\right\}\right\} \quad (71)$$

is the same for all $\{v_j\}_{j=2}^{K} \in \mathcal{B}_{1,2}$, and so on till $\{v_j\}_{j=2}^{K} \in \mathcal{B}_{N,2}$, but may be different for different $\mathcal{B}_{l,2}$, then

$$\Pr\left\{\bigcup_{i}\left\{\bigcup_{m=1}^{N}\left\{V_1(i) \in \mathcal{A}_{l,0}, \bigcap_{k=1}^{K-1}(V_1(i), V_{k+1}) \in \mathcal{A}_{m,k}, \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{m,k}\right\}\right\}\right\}$$

$$\geq \frac{1}{4}\min\left\{1, L_1 \Pr\left\{\bigcup_{m=1}^{N}\{V_1 \in \mathcal{B}_{m,1}\}\right\}, \Pr\left\{\bigcup_{m=1}^{N}\left\{\{V_j\}_{k=2}^{K} \in \mathcal{B}_{m,2}\right\}\right\},\right.$$

$$\left. L_1 \Pr\left\{\bigcup_{m=1}^{N}\left\{V_1 \in \mathcal{A}_{m,0}, \bigcap_{k=1}^{K-1}(V_1, V_{k+1}) \in \mathcal{A}_{m,k}, \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{m,k}\right\}\right\}\right\}. \quad (72)$$

Remark 1 Note that the number of sequences, K, can be arbitrarily large, and in particular, exponential, without affecting the tightness of the lower and upper bounds. Also, note that the above lemma can be easily generalized to the case where we have random sequences, $\{V_2(i)\}_{i=1}^{L_2}, \ldots, \{V_K(i)\}_{i=1}^{L_2}$, rather than single random variables $V_2, \ldots, V_K$, respectively.

Next, we apply Lemma 1 to the problem at hand. To this end, we choose the following parameters in accordance to the notations used in Lemma 1. Recall that we deal with:

$$\bigcup_{i=1}^{M_1-1}\bigcup_{\boldsymbol{l}}\left\{\begin{array}{c}\boldsymbol{X}_{1,i} \in \mathcal{A}_{\boldsymbol{l},0}, (\boldsymbol{X}_{1,i}, \boldsymbol{X}_{2,1}) \in \mathcal{A}_{\boldsymbol{l},1}, \ldots, (\boldsymbol{X}_{1,i}, \boldsymbol{X}_{2,M_2-1}) \in \mathcal{A}_{\boldsymbol{l},M_2-1} \\ \boldsymbol{X}_{2,1} \in \tilde{\mathcal{A}}_{\boldsymbol{l},1}, \ldots, \boldsymbol{X}_{2,M_2-1} \in \tilde{\mathcal{A}}_{\boldsymbol{l},M_2-1}\end{array}\right\}, \quad (73)$$

and in Lemma 1 we have considered:

$$\bigcup_{i}\bigcup_{m=1}^{N}\left\{\begin{array}{c} V_1(i)\in\mathcal{A}_{m,0},(V_1(i),V_2)\in\mathcal{A}_{m,1},\ldots,(V_1(i),V_K)\in\mathcal{A}_{m,K-1} \\ V_2\in\mathcal{G}_{m,2},\ldots,V_K\in\mathcal{G}_{m,K}\end{array}\right\}. \tag{74}$$

Thus, comparing (73) and (74), we readily notice to the following parallels:

- The numbers of events in the unions over i is $L_1 = M_1 - 1$. Also, we have $K = M_2$ independent random vectors, where $V_1\left(i\right) = \boldsymbol{X}_{1,i}$, $V_l(1) = \boldsymbol{X}_{2,l}$ for $2 \leq l \leq M_2 - 1$. Again, since $J = 1$, we have fixed the index of $V_l(1)$ to 1.
- We have:

1) $\mathcal{A}_{m,i} = \mathcal{A}_{\boldsymbol{l},i},$ for $0 \leq i \leq M_2 - 1,$
2) $\mathcal{G}_{m,i} = \tilde{\mathcal{A}}_{\boldsymbol{l},i-1},$ for $2 \leq i \leq M_2.$

These sets correspond to each of the typical sets $T(\tilde{Q}^0_{X_1X_{2,0}Y})$, $\left\{T(\tilde{Q}^k_{X_1X_2Y})\right\}_{k=1}^{M_2-1}$, $\left\{T(\hat{Q}^k_{X_{1,0}X_2Y})\right\}_{k=1}^{M_2-1}$. Also, the union over m corresponds to a union over $\boldsymbol{l}$, which as was mentioned before, is actually a union over a vector of types.

- According to (67) and (68) we need to define $\mathcal{B}_{m,1} = \mathcal{B}_1(\tilde{Q}^0_{X_1X_{2,0}Y}, \left\{\tilde{Q}^j_{X_1X_2Y}, \hat{Q}^j_{X_1X_2Y}\right\}_j)$ and $\mathcal{B}_{m,2} = \mathcal{B}_2(\tilde{Q}^0_{X_1X_{2,0}Y}, \left\{\tilde{Q}^j_{X_1X_2Y}, \hat{Q}^j_{X_1X_2Y}\right\}_j)$. Accordingly, by the definitions given in (67) and (68), we get

$$\mathcal{B}_{m,1} = \left\{\boldsymbol{x}_1 : \begin{array}{c} (\boldsymbol{x}_1,\boldsymbol{x}_{2,0},\boldsymbol{y})\in T(\tilde{Q}^0_{X_1X_{2,0}Y}), \\ \left\{(\boldsymbol{x}_1,\boldsymbol{x}_{2,j},\boldsymbol{y})\in T(\tilde{Q}^j_{X_1X_2Y})\right\}_{j=1}^{M_2-1}, \\ \left\{(\boldsymbol{x}_{1,0},\boldsymbol{x}_{2,j},\boldsymbol{y})\in T(\hat{Q}^j_{X_{1,0}X_2Y})\right\}_{j=1}^{M_2-1} \text{ for some } \left\{\boldsymbol{x}_{2,j}\right\}_j \end{array}\right\}, \tag{75}$$

and

$$\mathcal{B}_{m,2} = \left\{\left\{\boldsymbol{x}_{2,j}\right\}_{j\geq1} : \begin{array}{c} (\boldsymbol{x}_1,\boldsymbol{x}_{2,0},\boldsymbol{y})\in T(\tilde{Q}^0_{X_1X_{2,0}Y}), \\ \left\{(\boldsymbol{x}_1,\boldsymbol{x}_{2,j},\boldsymbol{y})\in T(\tilde{Q}^j_{X_1X_2Y})\right\}_{j=1}^{M_2-1}, \\ \left\{(\boldsymbol{x}_{1,0},\boldsymbol{x}_{2,j},\boldsymbol{y})\in T(\hat{Q}^j_{X_{1,0}X_2Y})\right\}_{j=1}^{M_2-1} \text{ for some } \boldsymbol{x}_1 \end{array}\right\}. \tag{76}$$

Finally, note that the requirements (70) and (71) in Lemma 1 hold. For example, the requirement in (70) means that the probability

$$\Pr\left\{\bigcup_{\left\{\tilde{Q}^j_{X_1X_2Y},\hat{Q}^j_{X_1X_2Y}\right\}_j\in T_I(Q_{X_{1,0}X_{2,0}Y})}\left\{\begin{array}{c} (\boldsymbol{x}_{1,1},\boldsymbol{x}_{2,0},\boldsymbol{y})\in T(\tilde{Q}^0_{X_1X_{2,0}Y}), \\ \left\{(\boldsymbol{x}_{1,1},\boldsymbol{X}_{2,j},\boldsymbol{y})\in T(\tilde{Q}^j_{X_1X_2Y})\right\}_{j=1}^{M_2-1}, \\ \left\{(\boldsymbol{x}_{1,0},\boldsymbol{X}_{2,j},\boldsymbol{y})\in T(\hat{Q}^j_{X_{1,0}X_2Y})\right\}_{j=1}^{M_2-1}\end{array}\right\}\right\} \tag{77}$$

is constant over $\mathcal{B}_{m,1}$ for every m (but may be different for different m). This is true because everything is expressed in terms of types. Indeed, if we fix m, then over $\mathcal{B}_{m,1}$, the first and third constraints in the event of (77) are held fixed, and the second constraint is also independent of the specific sequence $\boldsymbol{x}_{1,1}$ from $\mathcal{B}_{m,1}$ because the joint empirical distribution of $(\boldsymbol{x}_{1,1},\boldsymbol{y})$ is fixed to $\tilde{Q}^0_{X_1Y}$, and this type is consistent with the distribution $\tilde{Q}^j_{X_1X_2Y}$, which have, by construction, the same marginal of $(\boldsymbol{x}_{1,1},\boldsymbol{y})$. Due to the same reasoning, the probability

$$\Pr\left\{\bigcup_{\left\{\tilde{Q}^j_{X_1X_2Y},\hat{Q}^j_{X_1X_2Y}\right\}_j\in T_I(Q_{X_{1,0}X_{2,0}Y})}\left\{\begin{array}{c}(\boldsymbol{X}_{1,1},\boldsymbol{x}_{2,0},\boldsymbol{y})\in T(\tilde{Q}^0_{X_1X_{2,0}Y}),\\ \left\{(\boldsymbol{X}_{1,1},\boldsymbol{x}_{2,j},\boldsymbol{y})\in T(\tilde{Q}^j_{X_1X_2Y})\right\}_{j=1}^{M_2-1},\\ \left\{(\boldsymbol{x}_{1,0},\boldsymbol{x}_{2,j},\boldsymbol{y})\in T(\hat{Q}^j_{X_{1,0}X_2Y})\right\}_{j=1}^{M_2-1}\end{array}\right\}\right\} \tag{78}$$

is constant over $\mathcal{B}_{m,2}$ for every m (but may be different for different m).

Thus, invoking Lemma 1, we may write

$$\tilde{P}_e \triangleq \Pr\left[\bigcup_{i=1}^{M_1-1}\left\{\sum_{j=0}^{M_2-1}P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,i},\boldsymbol{X}_{2,j}\right)\geq\sum_{j=0}^{M_2-1}P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,0},\boldsymbol{X}_{2,j}\right)\right\}\middle|\mathcal{F}_0\right] \tag{79}$$

$$\begin{aligned}&\doteq\min\Bigg\{1,M_1\cdot\Pr\left[\bigcup_{\left\{\tilde{Q}^j_{X_1X_2Y},\hat{Q}^j_{X_1X_2Y}\right\}_j\in T_I(Q_{X_{1,0}X_{2,0}Y})}\boldsymbol{X}_{1,1}\in\mathcal{B}_1\left(\tilde{Q}^0_{X_1X_{2,0}Y},(\tilde{Q}^j_{X_1X_2Y},\hat{Q}^j_{X_1X_2Y})_j)\right)\right],\\ &\qquad\Pr\left[\bigcup_{\left\{\tilde{Q}^j_{X_1X_2Y},\hat{Q}^j_{X_1X_2Y}\right\}_j\in T_I(Q_{X_{1,0}X_{2,0}Y})}\{\boldsymbol{X}_{2,j}\}_{j\geq1}\in\mathcal{B}_2\left(\tilde{Q}^0_{X_1X_{2,0}Y},(\tilde{Q}^j_{X_1X_2Y},\hat{Q}^j_{X_1X_2Y})_j)\right)\right],\\ &M_1\cdot\Pr\left[\bigcup_{\left\{\tilde{Q}^j_{X_1X_2Y},\hat{Q}^j_{X_1X_2Y}\right\}_j\in T_I(Q_{X_{1,0}X_{2,0}Y})}\left\{\begin{array}{c}(\boldsymbol{X}_{1,1},\boldsymbol{x}_{2,0},\boldsymbol{y})\in T(\tilde{Q}^0_{X_1X_{2,0}Y}),\\ \left\{(\boldsymbol{X}_{1,1},\boldsymbol{X}_{2,j},\boldsymbol{y})\in T(\tilde{Q}^j_{X_1X_2Y})\right\}_{j=1}^{M_2-1},\\ \left\{(\boldsymbol{x}_{1,0},\boldsymbol{X}_{2,j},\boldsymbol{y})\in T(\hat{Q}^j_{X_{1,0}X_2Y})\right\}_{j=1}^{M_2-1}\end{array}\right\}\right]\Bigg\}\end{aligned} \tag{80}$$

where each of the probabilities at the r.h.s. of (80) are conditioned on $\mathcal{F}_0$. Therefore, we were able to simplify the problematic union over the codebook of the first user. Note, however, that we cannot (directly) apply here the method of types due to the fact that the union is over an exponential number of types, and thus a more refined analysis is needed. We start by analyzing the last term at the r.h.s. of (80). To this end, we will invoke the type enumeration method, but first, the main observation here is that similarly to the passage from (64) to (65), the last term at the r.h.s. of (80) can be rewritten as follows:

$$\Pr\left[\bigcup_{\left\{\tilde{Q}^j_{X_1X_2Y},\hat{Q}^j_{X_1X_2Y}\right\}_j\in T_I(Q_{X_{1,0}X_{2,0}Y})}\left\{\begin{array}{c}(\boldsymbol{X}_{1,1},\boldsymbol{x}_{2,0},\boldsymbol{y})\in T(\tilde{Q}^0_{X_1X_{2,0}Y}),\\ \left\{(\boldsymbol{X}_{1,1},\boldsymbol{X}_{2,j},\boldsymbol{y})\in T(\tilde{Q}^j_{X_1X_2Y})\right\}_{j=1}^{M_2-1},\\ \left\{(\boldsymbol{x}_{1,0},\boldsymbol{X}_{2,j},\boldsymbol{y})\in T(\hat{Q}^j_{X_{1,0}X_2Y})\right\}_{j=1}^{M_2-1}\end{array}\right\}\right]$$

$$= \Pr\left[\left\{\sum_{j=0}^{M_2-1}P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,1},\boldsymbol{X}_{2,j}\right)\geq\sum_{j=0}^{M_2-1}P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,0},\boldsymbol{X}_{2,j}\right)\right\}\middle|\mathcal{F}_0\right] \tag{81}$$

$$= \mathbb{E}\left\{\Pr\left[\left\{\sum_{j=0}^{M_2-1}P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,1},\boldsymbol{X}_{2,j}\right)\geq\sum_{j=0}^{M_2-1}P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,0},\boldsymbol{X}_{2,j}\right)\right\}\middle|\mathcal{F}_0,\boldsymbol{X}_{1,1}\right]\middle|\mathcal{F}_0\right\}. \tag{82}$$

That is, we returned back to the structure of the original probability, but now, without the union over the codebook of the first user. Note that conditioning on the random vector $\boldsymbol{X}_{1,1}$ in (82), is due to the fact that $\boldsymbol{X}_{1,1}$ is common to all the summands in the inner summation over the codebook of the second user. We next evaluate the exponential behavior of the probability in (82). For a given realization of $\boldsymbol{Y}=\boldsymbol{y}$, $\boldsymbol{X}_{1,0}=\boldsymbol{x}_{1,0}$, $\boldsymbol{X}_{1,1}=\boldsymbol{x}_{1,1}$, and $\boldsymbol{X}_{2,0}=\boldsymbol{x}_{2,0}$, let us define

$$s\triangleq\frac{1}{n}\log P\left(\boldsymbol{y}|\boldsymbol{x}_{1,0},\boldsymbol{x}_{2,0}\right), \tag{83}$$

and

$$r\triangleq\frac{1}{n}\log P\left(\boldsymbol{y}|\boldsymbol{x}_{1,1},\boldsymbol{x}_{2,0}\right). \tag{84}$$

For a given $(\boldsymbol{y},\boldsymbol{x}_{1,0},\boldsymbol{x}_{1,1},\boldsymbol{x}_{2,0})$, and a given joint probability distribution $Q_{X_1X_2Y}$ on $\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{Y}$, let $N_1\left(Q_{X_1X_2Y}\right)$ designate the number of codewords $\left\{\boldsymbol{X}_{2,j}\right\}_j$ (excluding $\boldsymbol{x}_{2,0}$) whose conditional empirical distribution with $\boldsymbol{y}$ and $\boldsymbol{x}_{1,1}$ is $Q_{X_1X_2Y}$, that is,

$$N_1\left(Q_{X_1X_2Y}\right)\triangleq\sum_{j=1}^{M_2-1}\mathcal{I}\left\{(\boldsymbol{x}_{1,1},\boldsymbol{X}_{2,j},\boldsymbol{y})\in T\left(Q_{X_1X_2Y}\right)\right\}, \tag{85}$$

and let $N_2\left(Q_{X_1X_2Y}\right)$ designate the number of codewords $\left\{\boldsymbol{X}_{2,j}\right\}_j$ (excluding $\boldsymbol{x}_{2,0}$) whose conditional empirical distribution with $\boldsymbol{y}$ and $\boldsymbol{x}_{1,0}$ is $Q_{X_1X_2Y}$, that is

$$N_2\left(Q_{X_1X_2Y}\right)\triangleq\sum_{j=1}^{M_2-1}\mathcal{I}\left\{(\boldsymbol{x}_{1,0},\boldsymbol{X}_{2,j},\boldsymbol{y})\in T\left(Q_{X_1X_2Y}\right)\right\}. \tag{86}$$

Also, recall that

$$f\left(Q_{X_1X_2Y}\right)=\frac{1}{n}\log P\left(\boldsymbol{y}|\boldsymbol{x}_1,\boldsymbol{x}_2\right) \tag{87}$$

$$=\sum_{(x_1,x_2,y)\in\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{Y}}Q_{X_1X_2Y}\left(x_1,x_2,y\right)\log W_{Y|X_1X_2}\left(y|x_1x_2\right) \tag{88}$$

where $Q_{X_1X_2Y}$ is understood to be the joint empirical distribution of $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) \in \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}^n$. Thus, in terms of the above notations, we may write:

$$\sum_{j=0}^{M_2-1} P\left(\boldsymbol{y}|\boldsymbol{x}_{1,1}, \boldsymbol{X}_{2,j}\right) = e^{nr} + \sum_{Q_{X_2|X_1Y} \in \mathcal{S}(Q_{X_1Y})} N_1\left(Q_{X_1X_2Y}\right) e^{nf(Q_{X_1X_2Y})} \tag{89}$$

$$\triangleq e^{nr} + \mathcal{N}_1(Q_{X_1Y}). \tag{90}$$

where for a given Q_{X_1Y}, $\mathcal{S}(Q_{X_1Y})$ is defined as the set of all distributions $\left\{Q_{X_2|X_1Y}\right\}$, such that $\sum_{(x_1,y)\in\mathcal{X}_1\times\mathcal{Y}} Q_{X_1Y}\left(x_1, y\right) Q_{X_2|X_1Y}\left(x_2|x_1, y\right) = P_{X_2}\left(x_2\right)$ for all $x_2 \in \mathcal{X}_2$. Similarly,

$$\sum_{j=0}^{M_2-1} P\left(\boldsymbol{y}|\boldsymbol{x}_{1,0}, \boldsymbol{X}_{2,j}\right) = e^{ns} + \sum_{Q_{X_2|X_{1,0}Y} \in \mathcal{S}(Q_{X_{1,0}Y})} N_2\left(Q_{X_{1,0}X_2Y}\right) e^{nf(Q_{X_{1,0}X_2Y})} \tag{91}$$

$$\triangleq e^{ns} + \mathcal{N}_2(Q_{X_{1,0}Y}). \tag{92}$$

where for a given $Q_{X_{1,0}Y}$, $\mathcal{S}(Q_{X_{1,0}Y})$ is defined as the set of all distributions $\left\{Q_{X_2|X_{1,0}Y}\right\}$, such that $\sum_{(x_1,y)\in\mathcal{X}_1\times\mathcal{Y}} Q_{X_{1,0}Y}\left(x_1, y\right) Q_{X_2|X_{1,0}Y}\left(x_2|x_1, y\right) = P_{X_2}\left(x_2\right)$ for all $x_2 \in \mathcal{X}_2$. For simplicity of notation, in the following, we use Q and $\tilde{Q}$ to denote $Q_{X_1X_2Y}$ and $Q_{X_{1,0}X_2Y}$, respectively. Therefore, with these definitions in mind, we wish to calculate (given $(\mathcal{F}_0, \boldsymbol{X}_{1,1})$)

$$\Pr\left[\sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}|\boldsymbol{x}_{1,1}, \boldsymbol{X}_{2,j}\right) \geq \sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}|\boldsymbol{x}_{1,0}, \boldsymbol{X}_{2,j}\right)\right] = \Pr\left[\mathcal{N}_1(Q_{X_1Y}) - \mathcal{N}_2(Q_{X_{1,0}Y}) \geq e^{ns} - e^{nr}\right]. \tag{93}$$

Let $\varepsilon > 0$ be arbitrarily small. Then,

$$\begin{aligned}
&\Pr\left[\mathcal{N}_1(Q_{X_1Y}) - \mathcal{N}_2(Q_{X_{1,0}Y}) \geq e^{ns} - e^{nr}\right] \\
&= \sum_i \Pr\left\{e^{ni\varepsilon} \leq \mathcal{N}_2(Q_{X_{1,0}Y}) \leq e^{n(i+1)\varepsilon}, \mathcal{N}_1(Q_{X_1Y}) - \mathcal{N}_2(Q_{X_{1,0}Y}) \geq e^{ns} - e^{nr}\right\} \\
&\leq \sum_i \Pr\left\{e^{ni\varepsilon} \leq \mathcal{N}_2(Q_{X_{1,0}Y}) \leq e^{(i+1)\varepsilon}, \mathcal{N}_1(Q_{X_1Y}) \geq e^{ni\varepsilon} + e^{ns} - e^{nr}\right\} &(94) \\
&= \sum_i \Pr\left\{e^{ni\varepsilon} \leq \mathcal{N}_2(Q_{X_{1,0}Y}) \leq e^{n(i+1)\varepsilon}\right\} \\
&\qquad \times \Pr\left\{\mathcal{N}_1(Q_{X_1Y}) \geq e^{ni\varepsilon} + e^{ns} - e^{nr} \middle| e^{ni\varepsilon} \leq \mathcal{N}_2(Q_{X_{1,0}Y}) \leq e^{n(i+1)\varepsilon}\right\} &(95)
\end{aligned}$$

where i ranges from $\frac{1}{n\epsilon}\log P(\boldsymbol{y}|\boldsymbol{x}_{1,0}, \boldsymbol{x}_{2,0})$ to R_2/ε. It is not difficult to show that can be show that (see, e.g., [16, p. 6028]):

$$\Pr\left\{e^{nt} \leq \mathcal{N}_2(Q_{X_{1,0}Y}) \leq e^{n(t+\varepsilon)}\right\} \doteq \begin{cases} 0 & t < t_0(Q_{X_{1,0}Y}) - \varepsilon \\ \exp\left[-nE(t, Q_{X_{1,0}Y})\right] & t \geq t_0(Q_{X_{1,0}Y}) \end{cases} \tag{96}$$

where

$$t_0(Q_{X_{1,0}Y}) \triangleq R_2 + \max_{\tilde{Q}\in\mathcal{S}(Q_{X_{1,0}Y}):\, I_{\tilde{Q}}(X_2;X_{1,0},Y)\leq R_2} \left[f(\tilde{Q}) - I_{\tilde{Q}}(X_2;X_{1,0},Y)\right], \tag{97}$$

and

$$E(t,Q_{X_{1,0}Y}) \triangleq \min\left\{\left[I_{\tilde{Q}}(X_2;X_{1,0},Y) - R_2\right]_+ :\ f(\tilde{Q}) + \left[R_2 - I_{\tilde{Q}}(X_2;X_{1,0},Y)\right]_+ \geq t\right\}. \tag{98}$$

Substituting the last result in (95), we get

$$\Pr\left[\mathcal{N}_1(Q_{X_1Y}) - \mathcal{N}_2(Q_{X_{1,0}Y}) \geq e^{ns} - e^{nr}\right]$$
$$\leq \sum_i \Pr\left\{e^{ni\varepsilon} \leq \mathcal{N}_2(Q_{X_{1,0}Y}) \leq e^{n(i+1)\varepsilon}, \mathcal{N}_1(Q_{X_1Y}) \geq e^{ni\varepsilon} + e^{ns} - e^{nr}\right\} \tag{99}$$
$$\doteq \sum_{i\geq t_0(Q_{X_{1,0}Y})/\varepsilon} \exp\left[-nE(i\varepsilon, Q_{X_{1,0}Y})\right]$$
$$\times \Pr\left\{\mathcal{N}_1(Q_{X_1Y}) \geq e^{ni\varepsilon} + e^{ns} - e^{nr} \middle| e^{ni\varepsilon} \leq \mathcal{N}_2(Q_{X_{1,0}Y}) \leq e^{n(i+1)\varepsilon}\right\}. \tag{100}$$

Next, we use the following lemma.

Lemma 2 Let $\{\mathcal{A}_k\}_{k\geq 0}$ and $\{\mathcal{B}_k\}_{k\geq 0}$ be a sequence of events, that may statistically depend each on another. If:

- The event $\mathcal{A}_0$ is an almost-sure event, i.e, $\Pr\{\mathcal{A}_0\} = 1$
- The probability $\Pr\{\mathcal{B}_k\}$ is monotonically decreasing as a function of k

Then,

$$\max_k \Pr\{\mathcal{A}_k \cap \mathcal{B}_k\} = \Pr\{\mathcal{B}_0\}. \tag{101}$$

Proof of Lemma 2: Note that:

$$\Pr\{\mathcal{A}_0 \cap \mathcal{B}_0\} \leq \max_k \Pr\{\mathcal{A}_k \cap \mathcal{B}_k\} \leq \max_k \Pr\{\mathcal{B}_k\} = \Pr\{\mathcal{B}_0\} = \Pr\{\mathcal{A}_0 \cap \mathcal{B}_0\} \tag{102}$$

where the first and second equalities follow from the second and first assumptions of this lemma, respectively. ■

We now apply Lemma 2 to (100), where $\mathcal{A}_k \triangleq \left\{e^{nk\varepsilon} \leq \mathcal{N}_2(Q_{X_{1,0}Y}) \leq e^{n(k+1)\varepsilon}\right\}$ and $\mathcal{B}_k \triangleq \left\{\mathcal{N}_1(Q_{X_1Y}) \geq e^{nk\varepsilon} + e^{ns} - e^{nr}\right\}$. Note that under this choice of $\mathcal{A}_k$ and $\mathcal{B}_k$, the assumptions of Lemma 2 hold, where $k=0$ in the lemma is replaced by $k = t_0(Q_{X_{1,0}Y})/\varepsilon$. Indeed, according to (96) the event $\mathcal{A}_{t_0}$ is an almost-sure event (the exponent $E(k\varepsilon, Q_{X_{1,0}Y})$ vanishes), and as shall be seen in the sequel, $\Pr\{\mathcal{B}_k\}$ is monotonically decreasing with k. Thus, applying Lemma 2, we conclude that the dominant contribution to the sum over i is due to the first term, $i = t_0(Q_{X_{1,0}Y})/\varepsilon$. Whence, using the above arguments and the fact that ε is arbitrarily small, we get by using standard large deviations techniques (see, e.g., [16, p. 6027])

$$\Pr\left[\mathcal{N}_1(Q_{X_1Y}) - \mathcal{N}_2(Q_{X_{1,0}Y}) \geq e^{ns} - e^{nr}\right] \doteq \Pr\left\{\mathcal{N}_1(Q_{X_1Y}) \geq e^{nt_0(Q_{X_{1,0}Y})} + e^{ns} - e^{nr}\right\} \tag{103}$$

$$\doteq \max_{Q\in\mathcal{S}(Q_{X_1Y})} \Pr\left\{N_1(Q) \geq e^{n\left[t_0(Q_{X_{1,0}Y}) - f(Q)\right]} + e^{n[s-f(Q)]} - e^{n[r-f(Q)]}\right\} \tag{104}$$

$$\doteq \max_{Q\in\mathcal{S}(Q_{X_1Y})} \begin{cases} 1 & r > \max\left[f(Q), t_0, s\right] \\ e^{-n[I_Q(X_2;X_1,Y)-R_2]_+} & r \leq \max\left[f(Q), t_0, s\right],\ Q \in \mathcal{L} \\ 0 & r \leq \max\left[f(Q), t_0, s\right],\ Q \in \mathcal{L}^c \end{cases} \tag{105}$$

$$= \exp\left\{-nE_1(Q_{X_1X_{2,0}Y}, Q_{X_{1,0}X_{2,0}Y})\right\} \tag{106}$$

where

$$\mathcal{L} \triangleq \left\{Q:\ \max\left[f(Q), t_0, s\right] - f(Q) \leq \left[R_2 - I_Q(X_2; X_1, Y)\right]_+\right\}. \tag{107}$$

Note that when $r > \max\left[f(Q), t_0, s\right]$, the r.h.s. term of the inequality in the probability in (104) is negative, and due to the fact that the enumerator is nonnegative, the overall probability is unity. Finally, we average over $\boldsymbol{X}_{1,1}$ given $\mathcal{F}_0$. Using the method of types we obtain

$$\mathbb{E}\left\{\Pr\left[\left\{\sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,1}, \boldsymbol{X}_{2,j}\right) \geq \sum_{j=0}^{M_2-1} P\left(\boldsymbol{Y}|\boldsymbol{X}_{1,0}, \boldsymbol{X}_{2,j}\right)\right\}\middle|\mathcal{F}_0, \boldsymbol{X}_{1,1}\right]\middle|\mathcal{F}_0\right\} \tag{108}$$

$$\doteq \exp\left\{-n \min_{Q_{X_1|X_{2,0}Y}\in\mathcal{S}(Q_{X_{1,0}X_{2,0}Y})} \left[I_Q(X_1; X_{2,0}, Y) + E_1(Q_{X_1X_{2,0}Y}, Q_{X_{1,0}X_{2,0}Y})\right]\right\} \tag{109}$$

$$\triangleq \exp\left\{-n\hat{E}_1(Q_{X_{1,0}X_{2,0}Y}, R_2)\right\}. \tag{110}$$

This completes the analysis of the last term at the r.h.s. of (80).

Next, we analyze the second and third terms at the r.h.s. of (80). Recall that the later is given by:

$$P_{e,3} \triangleq \Pr\left[\bigcup_{\left\{\tilde{Q}^j_{X_1X_2Y}, \hat{Q}^j_{X_1X_2Y}\right\}_j \in T_I(Q_{X_{1,0}X_{2,0}Y})} \{\boldsymbol{X}_{2,j}\}_{j\geq 1} \in \mathcal{B}_2\left(\tilde{Q}^0_{X_1X_{2,0}Y}, (\tilde{Q}^j_{X_1X_2Y}, \hat{Q}^j_{X_1X_2Y})_j)\right)\right]. \tag{111}$$

Accordingly, in the spirit of (82), we note that $P_{e,3}$ can be equivalently rewritten as:

$$P_{e,3} = \Pr\left[\bigcup_{Q_{X_1|X_{2,0}Y}} \sum_{j=0}^{M_2-1} P\left(\boldsymbol{y}|\boldsymbol{x}_{1,1}, \boldsymbol{X}_{2,j}\right) \geq\right.$$

$$\sum_{j=0}^{M_2-1} P\left(\boldsymbol{y}|\boldsymbol{x}_{1,0},\boldsymbol{X}_{2,j}\right), \text{ for some } \boldsymbol{x}_{1,1}\in T(Q_{X_1X_{2,0}Y})\Bigg|\mathcal{F}_0\Bigg]. \tag{112}$$

Note that in comparison to the probability that we have analyzed before, here $\boldsymbol{x}_{1,1}$ is some given sequence from a type that leads to erroneous decoding. Continuing, we may write

$$P_{e,3} \doteq \max_{Q_{X_1|X_{2,0}Y}\in\mathcal{S}(Q_{X_{1,0}X_{2,0}Y})} \Pr\left[\sum_{j=0}^{M_2-1} P\left(\boldsymbol{y}|\boldsymbol{x}_{1,1},\boldsymbol{X}_{2,j}\right) \geq \sum_{j=0}^{M_2-1} P\left(\boldsymbol{y}|\boldsymbol{x}_{1,0},\boldsymbol{X}_{2,j}\right)\Bigg|\mathcal{F}_0\right]. \tag{113}$$

However, the probability in (113) is exactly what we have already analyzed above, and thus we get

$$P_{e,3} \doteq \max_{Q_{X_1|X_{2,0}Y}\in\mathcal{S}(Q_{X_{1,0}X_{2,0}Y})} \exp\left\{-nE_1(Q_{X_1X_{2,0}Y},Q_{X_{1,0}X_{2,0}Y})\right\} \tag{114}$$

$$= \exp\left\{-n \min_{Q_{X_1|X_{2,0}Y}\in\mathcal{S}(Q_{X_{1,0}X_{2,0}Y})} \left[E_1(Q_{X_1X_{2,0}Y},Q_{X_{1,0}X_{2,0}Y})\right]\right\} \tag{115}$$

$$\triangleq \exp\left\{-n\hat{E}_2(Q_{X_{1,0}X_{2,0}Y},R_2)\right\}. \tag{116}$$

Note that the difference between $\hat{E}_1(Q_{X_{1,0}X_{2,0}Y},R_2)$ and $\hat{E}_2(Q_{X_{1,0}X_{2,0}Y},R_2)$ is the additional mutual information term, $I_Q(X_1;X_{2,0},Y)$, in $\hat{E}_1(Q_{X_{1,0}X_{2,0}Y},R_2)$, which is due to the averaging over $\boldsymbol{X}_{1,1}$. This completes the analysis of the third term at the r.h.s. of (80). Finally, recall that the second term at the r.h.s. of (80) is given by

$$A \triangleq M_1\cdot\Pr\left[\bigcup_{T_I(Q_{X_{1,0}X_{2,0}Y})} \boldsymbol{X}_{1,1}\in\mathcal{B}_1\left(\tilde{Q}^0_{X_1X_{2,0}Y},(\tilde{Q}^j_{X_1X_2Y},\hat{Q}^j_{X_1X_2Y})_j)\right)\right] \tag{117}$$

and is equivalent to

$$A = M_1\cdot\Pr\left[\bigcup_{T_I(Q_{X_{1,0}X_{2,0}Y})}\left\{\begin{array}{c}(\boldsymbol{X}_{1,1},\boldsymbol{x}_{2,0},\boldsymbol{y})\in T(\tilde{Q}^0_{X_1X_{2,0}Y}),\\ \left\{(\boldsymbol{X}_{1,1},\boldsymbol{x}_{2,j},\boldsymbol{y})\in T(\tilde{Q}^j_{X_1X_2Y})\right\}_{j=1}^{M_2-1}, \text{ for some } \{\boldsymbol{x}_{2,j}\}\\ \left\{(\boldsymbol{x}_{1,0},\boldsymbol{x}_{2,j},\boldsymbol{y})\in T(\hat{Q}^j_{X_{1,0}X_2Y})\right\}_{j=1}^{M_2-1}\end{array}\right\}\right]. \tag{118}$$

This term can be analyzed as before, but, we claim that it is actually larger than the fourth term at the r.h.s. of (80), and thus, essentially, does not affect the minimum in (80). Indeed, recall that the fourth term is given by

$$B \triangleq M_1\cdot\Pr\left[\bigcup_{T_I(Q_{X_{1,0}X_{2,0}Y})}\left\{\begin{array}{c}(\boldsymbol{X}_{1,1},\boldsymbol{x}_{2,0},\boldsymbol{y})\in T(\tilde{Q}^0_{X_1X_{2,0}Y}),\\ \left\{(\boldsymbol{X}_{1,1},\boldsymbol{X}_{2,j},\boldsymbol{y})\in T(\tilde{Q}^j_{X_1X_2Y})\right\}_{j=1}^{M_2-1},\\ \left\{(\boldsymbol{x}_{1,0},\boldsymbol{X}_{2,j},\boldsymbol{y})\in T(\hat{Q}^j_{X_{1,0}X_2Y})\right\}_{j=1}^{M_2-1}\end{array}\right\}\right], \tag{119}$$

and since the factor M_1 is common to both A and B, we just need to compare the probabilities in these terms. However, it is obvious that the probability term in B is smaller (in the exponential scale) than the probability in A, due to the fact that events in the former are contained in the events in the latter. Indeed, this is equivalent to comparing between $\Pr\{(Z_1, Z_2) \in \mathcal{Z}\}$ and $\Pr\{(Z_1, z_2) \in \mathcal{Z},\ \text{for some } z_2 \in \mathcal{Z}_2\}$, where Z_1 and Z_2 are random variables that are defined over the alphabets $\mathcal{Z}_1$ and $\mathcal{Z}_2$, respectively, and $\mathcal{Z} \subseteq \mathcal{Z}_1 \times \mathcal{Z}_2$. Let $\mathcal{V} \triangleq \tilde{\mathcal{V}} \times \mathcal{Z}_2$, in which

$$\tilde{\mathcal{V}} \triangleq \{z_1 \in \mathcal{Z}_1 :\ (z_1, z_2) \in \mathcal{Z},\ \text{for some } z_2 \in \mathcal{Z}_2\}. \tag{120}$$

Then, it is obvious that $\mathcal{Z} \subseteq \mathcal{V}$, and thus

$$\Pr\{(Z_1, Z_2) \in \mathcal{Z}\} = \sum_{(z_1,z_2)\in\mathcal{Z}} P(z_1, z_2) \leq \sum_{(z_1,z_2)\in\mathcal{V}} P(z_1, z_2) \tag{121}$$

$$= \sum_{z_1 \in \tilde{\mathcal{V}}} P(z_1) = \Pr\{(Z_1, z_2) \in \mathcal{Z},\ \text{for some } z_2\}. \tag{122}$$

Wrapping up, using (56), (80), and the last results, after averaging w.r.t. $\mathcal{F}_0$, we get

$$P_e \doteq \mathbb{E}\left\{\min\left\{1, e^{-n(\hat{E}_1(Q_{X_{1,0}X_{2,0}Y}, R_2) - R_1)}, e^{-n\hat{E}_2(Q_{X_{1,0}X_{2,0}Y}, R_2)}\right\}\right\} \tag{123}$$

$$= \mathbb{E}\left\{\min\left\{e^{-n\left[\hat{E}_1(Q_{X_{1,0}X_{2,0}Y}, R_2) - R_1\right]_+}, e^{-n\hat{E}_2(Q_{X_{1,0}X_{2,0}Y}, R_2)}\right\}\right\} \tag{124}$$

$$= \mathbb{E}\left\{\exp\left[-n\max\left\{\left[\hat{E}_1(Q_{X_{1,0}X_{2,0}Y}, R_2) - R_1\right]_+, \hat{E}_2(Q_{X_{1,0}X_{2,0}Y}, R_2)\right\}\right]\right\} \tag{125}$$

$$\doteq \exp\left\{-n\left[\min_{Q_{Y|X_{1,0}X_{2,0}}}\left[D(Q_{Y|X_{1,0}X_{2,0}}||W_{Y|X_{1,0}X_{2,0}}|P_{X_{1,0}} \times P_{X_{2,0}}) + E^*(Q, R_1, R_2)\right]\right]\right\} \tag{126}$$

where

$$E^*(Q, R_1, R_2) \triangleq \max\left\{\left[\hat{E}_1(Q_{X_{1,0}X_{2,0}Y}, R_2) - R_1\right]_+, \hat{E}_2(Q_{X_{1,0}X_{2,0}Y}, R_2)\right\}. \tag{127}$$

B. Proof of Theorem 2:

Without loss of generality, we assume throughout, that the transmitted codewords are $\boldsymbol{x}_{1,0}$ and $\boldsymbol{x}_{2,0}$ which correspond to $\boldsymbol{z}_{11,0}, \boldsymbol{z}_{12,0}, \boldsymbol{z}_{21,0}$ and $\boldsymbol{z}_{22,0}$, and due to the fact that we will analyze the first decoder, for convenience, we use $\boldsymbol{y}$ instead of $\boldsymbol{y}_1$. Here, we distinguish between several types of errors. Recall that the overall error probability is given by

$$P_e = \Pr\left\{(\hat{M}_{11}, \hat{M}_{12}, \hat{M}_{21}) \neq (0, 0, 0)\right\}, \tag{128}$$

so there are seven possible types of errors: $(\hat{M}_{11} \neq 0, \hat{M}_{12} = 0, \hat{M}_{21} = 0)$, $(\hat{M}_{11} = 0, \hat{M}_{12} \neq 0, \hat{M}_{21} = 0)$, $(\hat{M}_{11} = 0, \hat{M}_{12} = 0, \hat{M}_{21} \neq 0)$, $(\hat{M}_{11} \neq 0, \hat{M}_{12} \neq 0, \hat{M}_{21} = 0)$, $(\hat{M}_{11} \neq 0, \hat{M}_{12} = 0, \hat{M}_{21} \neq 0)$, $(\hat{M}_{11} = 0, \hat{M}_{12} \neq 0, \hat{M}_{21} \neq 0)$, and $(\hat{M}_{11} \neq 0, \hat{M}_{12} \neq 0, \hat{M}_{21} \neq 0)$. Obviously, the exponent of the overall error probability in (128) is given by the minimum between error exponents of each of the type of error individually. Accordingly, we start with analyzing the last error event, which is also the most involved one. For this event, the average probability of error, associated with the decoder (22), is given by

$$P_e^{(7)} \triangleq \Pr\left[\bigcup_{i=1}^{M_{11}-1}\bigcup_{j=1}^{M_{12}-1}\bigcup_{k=1}^{M_{21}-1}\left\{\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{ijk},\boldsymbol{Z}_{22,l})\geq\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_0,\boldsymbol{Z}_{22,l})\right\}\right] \tag{129}$$

$$= \mathbb{E}\left\{\Pr\left[\bigcup_{i=1}^{M_{11}-1}\bigcup_{j=1}^{M_{12}-1}\bigcup_{k=1}^{M_{21}-1}\left\{\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{ijk},\boldsymbol{Z}_{22,l})\geq\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_0,\boldsymbol{Z}_{22,l})\right\}\middle|\mathcal{F}_0\right]\right\} \tag{130}$$

where $\tilde{\boldsymbol{Z}}_{ijk} \triangleq (\boldsymbol{Z}_{11,i},\boldsymbol{Z}_{12,j},\boldsymbol{Z}_{21,k})$, $\tilde{\boldsymbol{Z}}_0 \triangleq (\boldsymbol{Z}_{11,0},\boldsymbol{Z}_{12,0},\boldsymbol{Z}_{21,0})$, and $\mathcal{F}_0 \triangleq (\tilde{\boldsymbol{Z}}_0,\boldsymbol{Z}_{22,0},\boldsymbol{Y})$. For simplicity of notation, in the following, we use the indexes $\{1,2,3,4\}$ instead of $\{11,12,21,22\}$. We will assess the exponential behavior of (130) in the same manner as we did for (56). Specifically, we start with expressing (130) in terms of types. First, for a given joint distribution $Q_{Z_1^4Y}$, we let

$$f(Q_{Z_1^4Y}) \triangleq \frac{1}{n}\log P\left(\boldsymbol{y}|\boldsymbol{x}_1(\boldsymbol{z}_1,\boldsymbol{z}_2),\boldsymbol{x}_2(\boldsymbol{z}_3,\boldsymbol{z}_4)\right). \tag{131}$$

Now, for a given joint type $Q_{Z_{1,0}^4Y}$ of the random vectors $\left(\boldsymbol{Z}_{1,0}^4,\boldsymbol{Y}\right)$, we define the set:

$$T_I(Q_{Z_{1,0}^4Y})\triangleq\left\{\tilde{Q}^0_{Z_1^3Z_{4,0}Y}\in\mathcal{S}_0,\left(\left\{\tilde{Q}^l_{Z_1^4Y}\right\}_{l=1}^{M_{22}-1},\left\{\hat{Q}^l_{Z_{1,0}^3Z_4Y}\right\}_{l=1}^{M_{22}-1}\right)\in\mathcal{S}_1:\right.$$
$$\left.e^{nf(\tilde{Q}^0_{Z_1^3Z_{4,0}Y})}+\sum_{l=1}^{M_{22}-1}\left[e^{nf(\tilde{Q}^l_{Z_1^4Y})}-e^{nf(\hat{Q}^l_{Z_{1,0}^3Z_4Y})}\right]\geq e^{nf(Q_{Z_{1,0}^4Y})}\right\} \tag{132}$$

where

$$\mathcal{S}_0(Q_{Z_{1,0}^4Y}) \triangleq \left\{\tilde{Q}^0_{Z_1^3Z_{4,0}Y}:\ \tilde{Q}^0_{Z_1^3Z_{4,0}}=P_{Z_1^4},\tilde{Q}^0_{Z_{4,0}Y}=Q_{Z_{4,0}Y}\right\}, \tag{133}$$

and

$$\mathcal{S}_1(Q_{Z_{1,0}^4Y}) \triangleq \left\{\left\{\tilde{Q}^l_{Z_1^4Y}\right\}_{l=1}^{M_{22}-1},\left\{\hat{Q}^l_{Z_{1,0}^3Z_4Y}\right\}_{l=1}^{M_{22}-1}:\ \tilde{Q}^l_{Z_1^4}=P_{Z_1^4},\tilde{Q}^l_Y=Q_Y,\right.$$
$$\hat{Q}^l_{Z_{1,0}^3Z_4}=P_{Z_1^4},\hat{Q}^l_{Z_{1,0}^3Y}=Q_{Z_{1,0}^3Y},\ \forall 1\leq l\leq M_{22}-1$$
$$\left.\tilde{Q}^l_{Z_4Y}=\hat{Q}^l_{Z_4Y},\tilde{Q}^l_{Z_1^3Y}=\tilde{Q}^m_{Z_1^3Y},\forall l,m\right\}. \tag{134}$$

Now, with these definitions, fixing $Q_{Z_{1,0}^4Y}$, it follows, by definition, that the error event

$$\bigcup_{i=1}^{M_{11}-1}\bigcup_{j=1}^{M_{12}-1}\bigcup_{k=1}^{M_{21}-1}\left\{\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{ijk},\boldsymbol{Z}_{4,l})\geq\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_0,\boldsymbol{Z}_{4,l})\right\} \tag{135}$$

can be rewritten, in terms of types, as follows:

$$\bigcup_{i=1}^{M_{11}-1}\bigcup_{j=1}^{M_{12}-1}\bigcup_{k=1}^{M_{21}-1}\bigcup_{T_I(Q_{Z_{1,0}^4Y})}\left\{\begin{array}{c}(\tilde{\boldsymbol{Z}}_{ijk},\boldsymbol{z}_{4,0},\boldsymbol{y})\in T(\tilde{Q}^0_{Z_1^3Z_{4,0}Y}),\\ \left\{(\tilde{\boldsymbol{Z}}_{ijk},\boldsymbol{Z}_{4,l},\boldsymbol{y})\in T(\tilde{Q}^l_{Z_1^4Y})\right\}_{l=1}^{M_{22}-1},\\ \left\{(\tilde{\boldsymbol{z}}_0,\boldsymbol{Z}_{4,l},\boldsymbol{y})\in T(\hat{Q}^l_{Z_{1,0}^3Z_4Y})\right\}_{l=1}^{M_{22}-1}\end{array}\right\}. \tag{136}$$

We wish to analyze the probability of (136), conditioned on $\mathcal{F}_0$. Note that the inner union in (136) is over vectors of types (an exponential number of them). Finally, for the sake of convenience, we simplify the notations of (136), and write it equivalently as

$$\bigcup_{i=1}^{M_{11}-1}\bigcup_{j=1}^{M_{12}-1}\bigcup_{k=1}^{M_{21}-1}\bigcup_{\boldsymbol{l}}\left\{\begin{array}{c}\tilde{\boldsymbol{Z}}_{ijk}\in\mathcal{A}_{\boldsymbol{l},0},\\ (\tilde{\boldsymbol{Z}}_{ijk},\boldsymbol{Z}_{4,m})\in\mathcal{A}_{\boldsymbol{l},m},\quad \text{for } m=1,\ldots,M_{22}-1\\ \boldsymbol{Z}_{4,m}\in\tilde{\mathcal{A}}_{\boldsymbol{l},m},\quad \text{for } m=1,\ldots,M_{22}-1\end{array}\right\}. \tag{137}$$

where, again, the index "$\boldsymbol{l}$" in the inner union runs over the combinations of types (namely, $\boldsymbol{l}=\left\{\tilde{Q}^l_{Z_1^4Y},\hat{Q}^l_{Z_{1,0}^3Z_4Y}\right\}_l$) that belong to $T_I(Q_{Z_{1,0}^4Y})$, and the various sets $\left\{\mathcal{A}_{\boldsymbol{l},j},\tilde{\mathcal{A}}_{\boldsymbol{l},j}\right\}_{\boldsymbol{l},j}$ correspond to the typical sets in (136) (recall that $(\boldsymbol{z}_{1,0}^4,\boldsymbol{y})$ are given in this stage). Next, as before, we derive tight lower and upper bounds on a generic probability which have the form of (137). In the following, we give a generalization of Lemma 1 to the probability of a union indexed by K values, which is stated without proof. For a given subset $\mathcal{J}=\left\{j_1,\ldots,k_{|\mathcal{J}|}\right\}$ of $\{1,\ldots,J\}$ we write $\boldsymbol{Z}_{\mathcal{J}}$ as a shorthand for $(Z_{j_1},\ldots,Z_{j_{|\mathcal{J}|}})$.

Lemma 3 Let $\{Z_1(i)\}_{i=1}^{N_1},\ldots,\{Z_J(i)\}_{i=1}^{N_J},\{V_1(i)\}_{i=1}^{N_{J+1}},\{V_2(i)\}_{i=1}^{N_{J+1}},\ldots,\{V_K(i)\}_{i=1}^{N_{J+1}}$ be independent sequences of independently and identically distributed (i.i.d.) random variables on the alphabets $\mathcal{Z}_1\times\ldots\times\mathcal{Z}_J\times\mathcal{V}_1\times\ldots\times\mathcal{V}_K$, respectively, with $Z_1(i)\sim P_{Z_1},\ldots,Z_J(i)\sim P_{Z_J},V_1(i)\sim P_{V_1},\ldots,V_K(i)\sim P_{V_K}$. Fix a sequence of sets $\{\mathcal{A}_{i,1}\}_{i=1}^N,\{\mathcal{A}_{i,2}\}_{i=1}^N,\ldots,\{\mathcal{A}_{i,K}\}_{i=1}^N$, where $\mathcal{A}_{i,j}\subseteq\mathcal{Z}_1\times\ldots\times\mathcal{Z}_J\times\mathcal{V}_j$, for $1\leq j\leq K$ and for all $1\leq i\leq N$. Also, fix a set $\{\mathcal{A}_{i,0}\}_{i=1}^N$ where $\mathcal{A}_{i,0}\subseteq\mathcal{Z}_1\times\ldots\times\mathcal{Z}_J$ for all $1\leq i\leq N$, and another sequence of sets $\{\mathcal{G}_{i,1}\}_{i=1}^N,\{\mathcal{G}_{i,2}\}_{i=1}^N,\ldots,\{\mathcal{G}_{i,K}\}_{i=1}^N$, where $\mathcal{G}_{i,j}\subseteq\mathcal{V}_j$, for $1\leq j\leq K$ and for all $1\leq i\leq N$. Let $\boldsymbol{U}=(Z_1,Z_2,\ldots,Z_J,U_{J+1})$ with $U_{J+1}\triangleq(V_1,\ldots,V_K)$. Finally, define

$$\mathcal{B}_{l,\mathcal{J}}\triangleq\left\{\boldsymbol{u}_{\mathcal{J}}:\ z_1^J\in\mathcal{A}_{l,0},\ \bigcap_{j=1}^K\left(z_1^J,v_j\right)\in\mathcal{A}_{l,j},\ \bigcap_{j=1}^K v_j\in\mathcal{G}_{l,j}\ \text{ for some } \boldsymbol{u}_{\mathcal{J}^c}\right\}, \tag{138}$$

for $l=1,2,\ldots,N$, and $\mathcal{Z}(i_1^J)=(Z_1(i_1),\ldots,Z_J(i_J))$. Then,

1) A general upper bound is given by (we denote $\mathcal{Z}(i_1^J) \triangleq (Z_1(i_1), \ldots, Z_J(i_J))$)

$$\Pr\left\{\bigcup_{i_1^J,j}\left\{\bigcup_{l=1}^{N}\left\{\mathcal{Z}(i_1^J)\in\mathcal{A}_{l,0},\ \bigcap_{k=1}^{K}\big(\mathcal{Z}(i_1^J),V_k(j)\big)\in\mathcal{A}_{l,k},\ \bigcap_{k=1}^{K}V_k(j)\in\mathcal{G}_{l,k}\right\}\right\}\right\}$$
$$\leq \min\left\{1, \min_{\mathcal{J}\subseteq\{1,\ldots,J+1\}\mathcal{J}\neq\emptyset}\left(\prod_{j\in\mathcal{J}}N_j\right)\Pr\left\{\bigcup_{l=1}^{N}\boldsymbol{U}_{\mathcal{J}}\in\mathcal{B}_{l,\mathcal{J}}\right\}\right\}. \tag{139}$$

2) If the above are independent sequences of pairwise independent and identically distributed random variables , and

$$\Pr\left\{\bigcup_{l=1}^{N}\left\{\mathcal{Z}(i_1^J)\in\mathcal{A}_{l,0},\ \bigcap_{k=1}^{K}\big(\mathcal{Z}(i_1^J),V_k(j)\big)\in\mathcal{A}_{l,k},\ \bigcap_{k=1}^{K}V_k(j)\in\mathcal{G}_{l,k}\right\}\middle|\boldsymbol{U}_{\mathcal{J}}=\boldsymbol{u}_{\mathcal{J}}\right\} \tag{140}$$

is the same for all $\boldsymbol{u}_{\mathcal{J}}\in\mathcal{B}_{1,\mathcal{J}}$, and for all $\boldsymbol{u}_{\mathcal{J}}\in\mathcal{B}_{2,\mathcal{J}}$, and so on till $\boldsymbol{u}_{\mathcal{J}}\in\mathcal{B}_{N,\mathcal{J}}$, but may be different for different $\mathcal{B}_{l,\mathcal{J}}$, for a given $\mathcal{J}\subseteq\{1,\ldots,J+1\}$, then

$$\Pr\left\{\bigcup_{i_1^J,j}\left\{\bigcup_{l=1}^{N}\left\{\mathcal{Z}(i_1^J)\in\mathcal{A}_{l,0},\ \bigcap_{k=1}^{K}\big(\mathcal{Z}(i_1^J),V_k(j)\big)\in\mathcal{A}_{l,k},\ \bigcap_{k=1}^{K}V_k(j)\in\mathcal{G}_{l,k}\right\}\right\}\right\}$$
$$\geq 2^{-(J+1)}\min\left\{1, \min_{\mathcal{J}\subseteq\{1,\ldots,J+1\}\mathcal{J}\neq\emptyset}\left(\prod_{j\in\mathcal{J}}N_j\right)\Pr\left\{\bigcup_{l=1}^{N}\boldsymbol{U}_{\mathcal{J}}\in\mathcal{B}_{l,\mathcal{J}}\right\}\right\}. \tag{141}$$

Applying Lemma 3 on (136) (or, (137)) we obtain

$$\Pr\left\{\bigcup_{i=1}^{M_{11}-1}\bigcup_{j=1}^{M_{12}-1}\bigcup_{k=1}^{M_{21}-1}\left\{\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{ijk},\boldsymbol{Z}_{4,l})\geq\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_0,\boldsymbol{Z}_{4,l})\right\}\middle|\mathcal{F}_0\right\}$$
$$\doteq \min\left\{1, \min_{\mathcal{J}\subseteq\{1,\ldots,4\}\mathcal{J}\neq\emptyset}\left(\prod_{j\in\mathcal{J}}N_j\right)\Pr\left\{\bigcup_{\boldsymbol{l}}\boldsymbol{U}_{\mathcal{J}}\in\mathcal{B}_{\boldsymbol{l},\mathcal{J}}\right\}\right\} \tag{142}$$

where $N_1=M_{11}, N_2=M_{12}, N_3=M_{21}, N_4=1$, and

$$\boldsymbol{U}=(\boldsymbol{Z}_{11},\boldsymbol{Z}_{12},\boldsymbol{Z}_{21},\boldsymbol{U}_4) \tag{143}$$

in which $\boldsymbol{U}_4=(\boldsymbol{Z}_{4,1},\ldots,\boldsymbol{Z}_{4,M_{22}-1})$, and

$$\mathcal{B}_{\boldsymbol{l},\mathcal{J}}=\left\{\begin{array}{c}(\tilde{\boldsymbol{z}}_{ijk},\boldsymbol{z}_{4,0},\boldsymbol{y})\in T(\tilde{Q}^0_{X_1X_{2,0}Y}),\\ \boldsymbol{u}_{\mathcal{J}}:\ \left\{(\tilde{\boldsymbol{z}}_{ijk},\boldsymbol{z}_{4,l},\boldsymbol{y})\in T(\tilde{Q}^l_{Z_1^4Y})\right\}_{l=1}^{M_{22}-1},\\ \left\{(\tilde{\boldsymbol{z}}_0,\boldsymbol{z}_{4,l},\boldsymbol{y})\in T(\hat{Q}^j_{X_{1,0}X_2Y})\right\}_{l=1}^{M_{22}-1},\ \text{for some }\boldsymbol{u}_{\mathcal{J}^c}\end{array}\right\}. \tag{144}$$

Now, the various possibilities for the set $\mathcal{J}$ are:

$$\left\{\begin{array}{c} 1;2;3;4; \\ 12;13;14;23;24;34; \\ 123;124;134;234; \\ 1234 \end{array}\right\}, \tag{145}$$

that is, we have 15 possibilities. Now, we claim that possibilities $\{1,2,3,12,13,23,123\}$ do not affect the outer minimum in (142), and so we left with possibilities $\{4,14,24,34,124,134,234,1234\}$. This is due to the same reasoning used in (122) for the second term at the r.h.s. of (80). Indeed, note that possibilities $\{1,2,3\}$ do not affect due to possibilities $14,24,34$, respectively. Indeed, the multiplicative factor for each of the pairs $((1,14)$, $(2,24)$, and $(3,34))$ is the same, but the respective probabilities in (142) are smaller for $14,24,34$. Similarly, possibilities $\{12,13,23,123\}$ do not affect due to possibilities $124,134,234,1234$, respectively.

In the following, we analyze the "surviving" terms. For example, the term that corresponds to possibility 1234, is given by

$$P_{e,1234} \triangleq M_{11}M_{12}M_{21}\Pr\left\{\bigcup_{\boldsymbol{l}}\boldsymbol{U}\in\mathcal{B}_{\boldsymbol{l},1234}\right\}, \tag{146}$$

which can be rewritten as

$$\begin{aligned} &M_{11}M_{12}M_{21}\Pr\left\{\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{111},\boldsymbol{Z}_{4,l})\geq\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{0},\boldsymbol{Z}_{4,l})\middle|\mathcal{F}_0\right\} \\ &=M_{11}M_{12}M_{21}\mathbb{E}\left\{\Pr\left\{\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{111},\boldsymbol{Z}_{4,l})\geq\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{0},\boldsymbol{Z}_{4,l})\middle|\mathcal{F}_0,\tilde{\boldsymbol{Z}}_{111}\right\}\middle|\mathcal{F}_0\right\}. \end{aligned} \tag{147}$$

But this has the same form of the probability in (108), which we already analyzed. Accordingly, we obtain:

$$\mathbb{E}\left\{\Pr\left\{\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{111},\boldsymbol{Z}_{4,l})\geq\sum_{l=0}^{M_{22}-1}P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{0},\boldsymbol{Z}_{4,l})\middle|\mathcal{F}_0,\tilde{\boldsymbol{Z}}_{111}\right\}\middle|\mathcal{F}_0\right\} \tag{148}$$

$$\doteq\exp\left\{-n\min_{Q_{Z_1^3|Z_{4,0}Y}\in\mathcal{S}(Q_{Z_{1,0}^4Y})}\left[I_Q(Z_1^3;Z_{4,0},Y)+E^{(7)}(Q_{Z_1^3Z_{4,0}Y},Q_{Z_{1,0}^4Y})\right]\right\} \tag{149}$$

$$\triangleq\exp\left\{-n\hat{E}_7^{(7)}(Q_{Z_{1,0}^4Y},R_{22})\right\}. \tag{150}$$

where

$$E^{(7)}(Q_{Z_1^3Z_{4,0}Y},Q_{Z_{1,0}^4Y})\triangleq\min_{\hat{Q}:\,\hat{Q}_{Z_1^3Y}=Q_{Z_1^3Y},\,\hat{Q}\in\mathcal{L}(Q_{Z_1^3Z_{4,0}Y},Q_{Z_{1,0}^4Y})}\left[I_{\hat{Q}}(Z_4;Z_1^3,Y)-R_{22}\right]_+, \tag{151}$$

in which

$$\mathcal{L}(Q_{Z_1^3 Z_{4,0} Y}, Q_{Z_{1,0}^4 Y}) \triangleq \Big\{ \hat{Q} : \max\Big[f(\hat{Q}), t_0(Q_{Z_{1,0}^4 Y}), f(Q_{Z_{1,0}^4 Y}) \Big] - f(\hat{Q}) \leq \Big[R_{22} - I_{\hat{Q}}(Z_4; Z_1^3, Y) \Big]_+ ,$$
$$f(Q_{Z_1^3 Z_{4,0} Y}) \leq \max\Big[f(\hat{Q}), t_0(Q_{Z_{1,0}^4 Y}), f(Q_{Z_{1,0}^4 Y}) \Big] \Big\}, \tag{152}$$

and

$$t_0(Q_{Z_{1,0}^4 Y}) \triangleq R_{22} + \max_{\hat{Q}:\, \hat{Q}_{Z_1^3 Y} = Q_{Z_{1,0}^3 Y},\, I_{\hat{Q}}(Z_4; Z_1^3, Y) \leq R_{22}} \Big[f(\hat{Q}) - I_{\hat{Q}}(Z_4; Z_1^3, Y) \Big]. \tag{153}$$

The other terms are handled in a similar fashion. Specifically, let $\hat{\boldsymbol{Z}} \triangleq \{Z_1, Z_2, Z_3\}$, and define the sets $\mathcal{U} = \{1,2,3,12,13,23,123\}$, and $\tilde{\mathcal{U}} = \{14,24,34,124,134,234,1234\}$. Then, define for any $u \in \{1,2,\ldots,7\}$:

$$P_{e,u}^{(7)} \triangleq M_{\mathcal{U}(u)} \cdot \Pr\Big\{ \bigcup_{\boldsymbol{l}} \boldsymbol{U}_{\mathcal{U}(u)} \in \mathcal{B}_{\boldsymbol{l},\mathcal{U}(u)} \Big\}, \tag{154}$$

where

$$M_{\mathcal{U}(1)} \triangleq M_{11};\ M_{\mathcal{U}(2)} = M_{12};\ M_{\mathcal{U}(3)} = M_{21}; M_{\mathcal{U}(4)} = M_{11}M_{12};$$
$$M_{\mathcal{U}(5)} \triangleq M_{11}M_{21};\ M_{\mathcal{U}(6)} = M_{12}M_{21}\ M_{\mathcal{U}(7)} = M_{11}M_{12}M_{21}. \tag{155}$$

Accordingly, we have

$$P_{e,u}^{(7)} \doteq \exp\Big\{ -n \min_{Q_{Z_1^3|Z_{4,0}Y} \in \mathcal{S}(Q_{Z_{1,0}^4 Y})} \Big[I_Q(\hat{\boldsymbol{Z}}_{\mathcal{U}(u)}; Z_{4,0}, Y | \hat{\boldsymbol{Z}}_{123\setminus\mathcal{U}(u)}) + E^{(7)}(Q_{Z_1^3 Z_{4,0} Y}, Q_{Z_{1,0}^4 Y}) \Big] \Big\}$$
$$\triangleq \exp\Big\{ -n \hat{E}_u^{(7)}(Q_{Z_{1,0}^4 Y}, R_{22}) \Big\}. \tag{156}$$

Finally, for possability $\{4\}$, we have

$$P_{e,8}^{(7)} \doteq \exp\Big\{ -n \min_{Q_{Z_1^3|Z_{4,0}Y} \in \mathcal{S}(Q_{Z_{1,0}^4 Y})} E^{(7)}(Q_{Z_1^3 Z_{4,0} Y}, Q_{Z_{1,0}^4 Y}) \Big\}$$
$$\triangleq \exp\Big\{ -n \hat{E}_8(Q_{Z_{1,0}^4 Y}, R_{22}) \Big\}. \tag{157}$$

Wrapping up, after averaging w.r.t. $\mathcal{F}_0$, we get

$$P_e^{(7)} \doteq \mathbb{E}\Big\{ \min\Big\{ 1, \min_{u \in \{1:7\}} e^{-n\left[\hat{E}_u^{(7)}(Q_{Z_{1,0}^4 Y}, R_{22}) - n^{-1}\log M_{\tilde{\mathcal{U}}(u)}\right]}, e^{-n\hat{E}_8^{(7)}(Q_{Z_{1,0}^4 Y}, R_{22})} \Big\} \Big\} \tag{158}$$

$$= \mathbb{E}\Big\{ \min\Big\{ \min_{u \in \{1:7\}} e^{-n\left[\hat{E}_u^{(7)}(Q_{Z_{1,0}^4 Y}, R_{22}) - n^{-1}\log M_{\tilde{\mathcal{U}}(u)}\right]_+}, e^{-n\hat{E}_8^{(7)}(Q_{Z_{1,0}^4 Y}, R_{22})} \Big\} \Big\} \tag{159}$$

$$= \mathbb{E}\Big\{ \exp\Big[-n \max\Big\{ \max_u \Big[\hat{E}_u^{(7)}(Q_{Z_{1,0}^4 Y}, R_{22}) - \frac{1}{n}\log M_{\tilde{\mathcal{U}}(u)} \Big]_+, \hat{E}_8^{(7)}(Q_{Z_{1,0}^4 Y}, R_{22}) \Big\} \Big] \Big\} \tag{160}$$

$$\doteq \exp\left\{-n\left[\min_{Q_{Y|Z_{1,0}^4}}\left[D(Q_{Y|Z_{1,0}^4}||W_{Y|Z_{1,0}^4}|P_{Z_{1,0}^4}) + E_{\text{HK}}^{(7)}(Q_{Z_{1,0}^4 Y}, R_1, R_2)\right]\right]\right\} \tag{161}$$

where

$$E_{\text{HK}}^{(7)}(Q_{Z_{1,0}^4 Y}, R_1, R_2) \triangleq \max\left\{\max_{u\in\{1:7\}}\left[\hat{E}_u^{(7)}(Q_{Z_{1,0}^4 Y}, R_{22}) - \mathcal{R}_u\right]_+, \hat{E}_8^{(7)}(Q_{Z_{1,0}^4 Y}, R_{22})\right\}, \tag{162}$$

and $\mathcal{R}_u$ for $u = 1, 2, \ldots, 7$ is defined in (29).

This concludes the analysis of the error event $(\hat{M}_{11} \neq 0, \hat{M}_{12} \neq 0, \hat{M}_{21} \neq 0)$ in (128). The other types of errors are analyzed in a similar fashion. Indeed, for $(\hat{M}_{11} \neq 0, \hat{M}_{12} = 0, \hat{M}_{21} = 0)$, the average probability of error, associated with the decoder given in (22), is given by:

$$P_e^{(1)} = \Pr\left[\bigcup_{i=1}^{M_{11}-1}\left\{\sum_{l=0}^{M_{22}-1} P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{i00}, \boldsymbol{Z}_{22,l}) \geq \sum_{l=0}^{M_{22}-1} P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_0, \boldsymbol{Z}_{22,l})\right\}\right] \tag{163}$$

$$= \mathbb{E}\left\{\Pr\left[\bigcup_{i=1}^{M_{11}-1}\left\{\sum_{l=0}^{M_{22}-1} P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_{i00}, \boldsymbol{Z}_{22,l}) \geq \sum_{l=0}^{M_{22}-1} P(\boldsymbol{Y}|\tilde{\boldsymbol{Z}}_0, \boldsymbol{Z}_{22,l})\right\}\middle| \mathcal{F}_0\right]\right\} \tag{164}$$

where $\mathcal{F}_0 \triangleq (\tilde{\boldsymbol{Z}}_0, \boldsymbol{Z}_{22,0}, \boldsymbol{Y})$. Thus, due to the fact that $(\boldsymbol{Z}_{12,0}, \boldsymbol{Z}_{21,0})$ are now fixed, they play a same role as $\boldsymbol{Y}$ and $\boldsymbol{Z}_{22,0}$. Accordingly, we have

$$P_e^{(1)} \doteq \exp\left\{-n\left[\min_{Q_{Y|Z_{1,0}^4}}\left[D(Q_{Y|Z_{1,0}^4}||W_{Y|Z_{1,0}^4}|P_{Z_{1,0}^4}) + E_{\text{HK}}^{(1)}(Q_{Z_{1,0}^4 Y}, R_1, R_2)\right]\right]\right\} \tag{165}$$

where

$$E_{\text{HK}}^{(1)}(Q_{Z_{1,0}^4 Y}, R_1, R_2) \triangleq \max\left\{\left[\hat{E}^{(1)}(Q_{Z_{1,0}^4 Y}, R_{22}) - \mathcal{R}_1\right]_+, \hat{E}_8^{(1)}(Q_{Z_{1,0}^4 Y}, R_{22})\right\}, \tag{166}$$

and

$$\hat{E}^{(1)}(Q_{Z_{1,0}^4 Y}, R_{22}) = \min_{Q_{Z_1|Z_{2,0}^4 Y}\in\mathcal{S}(Q_{Z_{1,0}^4 Y})}\left[I_Q(Z_1; Z_{2,0}^4, Y) + E^{(1)}(Q_{Z_1 Z_{2,0}^4 Y}, Q_{Z_{1,0}^4 Y})\right], \tag{167}$$

and

$$\hat{E}_8^{(1)}(Q_{Z_{1,0}^4 Y}, R_{22}) = \min_{Q_{Z_1|Z_{2,0}^4 Y}\in\mathcal{S}(Q_{Z_{1,0}^4 Y})} E^{(1)}(Q_{Z_1 Z_{2,0}^4 Y}, Q_{Z_{1,0}^4 Y}), \tag{168}$$

where

$$E^{(1)}(Q_{Z_1 Z_{2,0}^4 Y}, Q_{Z_{1,0}^4 Y}) \triangleq \min_{\hat{Q}:\, \hat{Q}_{Z_1^3 Y} = Q_{Z_1 Z_{2,0}^3 Y},\, \hat{Q}\in\mathcal{L}(Q_{Z_1 Z_{2,0}^4 Y}, Q_{Z_{1,0}^4 Y})}\left[I_{\hat{Q}}(Z_4; Z_1, Z_{2,0}^3, Y) - R_{22}\right]_+. \tag{169}$$

In a similar manner, one obtains error exponents of $P_e^{(2)}$ and $P_e^{(3)}$, corresponding to $(\hat{M}_{11} = 0, \hat{M}_{12} \neq 0, \hat{M}_{21} = 0)$ and $(\hat{M}_{11} = 0, \hat{M}_{12} = 0, \hat{M}_{21} \neq 0)$, respectively. Indeed, $P_e^{(2)}$ is obtained by replacing the role of Z_1 with Z_2 and $\mathcal{R}_1$ with $\mathcal{R}_2$, in $P_e^{(1)}$, and $P_e^{(3)}$ is obtained by replacing the role of Z_1 with Z_3 and $\mathcal{R}_1$ with $\mathcal{R}_3$, in $P_e^{(1)}$. Similarly, $P_e^{(4)}$, corresponding to $(\hat{M}_{11} \neq 0, \hat{M}_{12} \neq 0, \hat{M}_{21} = 0)$, is given by:

$$P_e^{(4)} \doteq \exp\left\{-n\left[\min_{Q_{Y|Z_{1,0}^4}}\left[D(Q_{Y|Z_{1,0}^4}||W_{Y|Z_{1,0}^4}|P_{Z_{1,0}^4}) + E_{\text{HK}}^{(4)}(Q_{Z_{1,0}^4 Y}, R_1, R_2)\right]\right]\right\} \tag{170}$$

where

$$E_{\text{HK}}^{(4)}(Q_{Z_{1,0}^4 Y}, R_1, R_2) \triangleq \max\left\{\max_{u\in\{1,2,4\}}\left[\hat{E}_u^{(4)}(Q_{Z_{1,0}^4 Y}, R_{22}) - \mathcal{R}_u\right]_+, \hat{E}_8^{(4)}(Q_{Z_{1,0}^4 Y}, R_{22})\right\}, \tag{171}$$

and

$$\hat{E}_u^{(4)}(Q_{Z_{1,0}^4 Y}, R_{22}) = \min_{Q_{Z_1^2|Z_{3,0}^4 Y}\in\mathcal{S}(Q_{Z_{1,0}^4 Y})}\left[I_Q(\hat{\boldsymbol{Z}}_{\mathcal{U}(u)}; Z_{3,0}^4, Y|\hat{\boldsymbol{Z}}_{12\setminus\mathcal{U}(u)}) + E^{(4)}(Q_{Z_1^2 Z_{3,0}^4 Y}, Q_{Z_{1,0}^4 Y})\right], \tag{172}$$

and

$$E^{(4)}(Q_{Z_1^2 Z_{3,0}^4 Y}, Q_{Z_{1,0}^4 Y}) \triangleq \min_{\hat{Q}:\, \hat{Q}_{Z_1^3 Y} = Q_{Z_1^2 Z_{3,0} Y},\, \hat{Q}\in\mathcal{L}(Q_{Z_1^2 Z_{3,0}^4 Y}, Q_{Z_{1,0}^4 Y})}\left[I_{\hat{Q}}(Z_4; Z_1^2, Z_{3,0}, Y) - R_{22}\right]_+. \tag{173}$$

Finally, in a similar fashion, we can obtain the error exponents of $P_e^{(5)}$ and $P_e^{(6)}$, corresponding to $(\hat{M}_{11} \neq 0, \hat{M}_{12} = 0, \hat{M}_{21} \neq 0)$ and $(\hat{M}_{11} = 0, \hat{M}_{12} \neq 0, \hat{M}_{21} \neq 0)$, respectively. For $P_e^{(5)}$ we just need to replace the role of Z_2 with Z_3, and the minimization in (171) is over the indexes $\{1,3,5\}$, and $P_e^{(6)}$ is obtained by replacing the role of Z_1 with Z_3, and the minimization in (171) is over the indexes $\{2,3,6\}$.

APPENDIX A
PROOF OF LEMMA 1

In order to prove Lemma 1, we feel that it is more convenient and deductive to prove first a simpler version of it. To assist the reader, the road in proving Lemma 1 is as follows: we first state and prove Lemma 4, and then using this Lemma we state and prove Lemma 5, which is a special case of Lemma 1. Then, we prove Lemma 6, which is a generalization of Lemma 4. Finally, using Lemma 6 we eventually prove Lemma 1. In the following, we actually prove a generalized version of Lemma 1, where we consider random sequences, $\{V_2(i)\}_{i=1}^{L_2}, \ldots, \{V_K(i)\}_{i=1}^{L_2}$, rather than single random variables $V_2, \ldots, V_K$. Lemma 1 is then obtained on substituting $L_2 = 1$. We start with the following result which can be thought of as an extension of [12, Lemma 2].

Lemma 4 Let $\{V_1(i)\}_{i=1}^{L_1}$, $\{V_2(i)\}_{i=1}^{L_2}$, and $\{V_3(i)\}_{i=1}^{L_2}$ be independent sequences of i.i.d. random variables on the alphabets $\mathcal{V}_1 \times \mathcal{V}_2 \times \mathcal{V}_3$, respectively, with $V_1(i) \sim P_{V_1}$, $V_2(i) \sim P_{V_2}$, and $V_3(i) \sim P_{V_3}$.

For any sequence of sets $\{\mathcal{A}_{i,1}\}_{i=1}^{N}$ and $\{\mathcal{A}_{i,2}\}_{i=1}^{N}$ such that $\mathcal{A}_{i,1} \subseteq \mathcal{V}_1 \times \mathcal{V}_2$ and $\mathcal{A}_{i,2} \subseteq \mathcal{V}_1 \times \mathcal{V}_3$ for all $1 \leq i \leq N$, we have

$$\begin{aligned}
&\Pr\left\{\bigcup_{i,j}\left\{\bigcup_{l=1}^{N}\left\{\left(V_1(i), V_2(j)\right) \in \mathcal{A}_{l,1},\ \left(V_1(i), V_3(j)\right) \in \mathcal{A}_{l,2}\right\}\right\}\right\} \\
&\leq \min\left\{1, L_1 \mathbb{E}\left[\min\left\{1, L_2 \Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1, V_2) \in \mathcal{A}_{l,1},\ (V_1, V_3) \in \mathcal{A}_{l,2}\right\} \middle| V_1\right\}\right\}\right],\right. \\
&\qquad \left. L_2 \mathbb{E}\left[\min\left\{1, L_2 \Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1, V_2) \in \mathcal{A}_{l,1},\ (V_1, V_3) \in \mathcal{A}_{l,2}\right\} \middle| V_2, V_3\right\}\right\}\right]\right\} \qquad \text{(A.1)}
\end{aligned}$$

with $(V_1, V_2, V_3) \sim P_{V_1} \times P_{V_2} \times P_{V_3}$. Also, if $\{V_1(i)\}_{i=1}^{L_1}$ are pairwise independent, $\{V_2(i)\}_{i=1}^{L_2}$ are pairwise independent, $\{V_3(i)\}_{i=1}^{L_2}$ are pairwise independent, then

$$\begin{aligned}
&\Pr\left\{\bigcup_{i,j}\left\{\bigcup_{l=1}^{N}\left\{\left(V_1(i), V_2(j)\right) \in \mathcal{A}_{l,1},\ \left(V_1(i), V_3(j)\right) \in \mathcal{A}_{l,2}\right\}\right\}\right\} \\
&\geq \frac{1}{4}\min\left\{1, L_1 \frac{\Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1, V_2) \in \mathcal{A}_{l,1},\ (V_1, V_3) \in \mathcal{A}_{l,2}\right\}\right\}^2}{\Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1, V_2) \in \mathcal{A}_{l,1}, (V_1, V_3) \in \mathcal{A}_{l,2}\right\} \cap \bigcup_{l=1}^{N}\left\{(V_1, V_2') \in \mathcal{A}_{l,1}, (V_1, V_3') \in \mathcal{A}_{l,2}\right\}\right\}},\right. \\
&\qquad L_2 \frac{\Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1, V_2) \in \mathcal{A}_{l,1},\ (V_1, V_3) \in \mathcal{A}_{l,2}\right\}\right\}^2}{\Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1, V_2) \in \mathcal{A}_{l,1}, (V_1, V_3) \in \mathcal{A}_{l,2}\right\} \cap \bigcup_{l=1}^{N}\left\{(V_1', V_2) \in \mathcal{A}_{l,1}, (V_1', V_3) \in \mathcal{A}_{l,2}\right\}\right\}}, \\
&\qquad \left. L_1 L_2 \Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1, V_2) \in \mathcal{A}_{l,1},\ (V_1, V_3) \in \mathcal{A}_{l,2}\right\}\right\}\right\} \qquad \text{(A.2)}
\end{aligned}$$

where $(V_1, V_1', V_2, V_2', V_3, V_3') \sim P_{V_1}(v_1) \times P_{V_1}(v_1') \times P_{V_2}(v_2) \times P_{V_2}(v_2') \times P_{V_3}(v_3) \times P_{V_3}(v_3')$.

Proof of Lemma 4: Starting with the upper bound, the second term in (A.1) follows by first applying the union bound over i

$$\begin{aligned}
&\Pr\left\{\bigcup_{i,j}\left\{\bigcup_{l=1}^{N}\left\{\left(V_1(i), V_2(j)\right) \in \mathcal{A}_{l,1},\ \left(V_1(i), V_3(j)\right) \in \mathcal{A}_{l,2}\right\}\right\}\right\} \\
&\leq L_1 \Pr\left\{\bigcup_{j}\left\{\bigcup_{l=1}^{N}\left\{\left(V_1, V_2(j)\right) \in \mathcal{A}_{l,1},\ \left(V_1, V_3(j)\right) \in \mathcal{A}_{l,2}\right\}\right\}\right\} \qquad \text{(A.3)} \\
&\leq L_1 \mathbb{E}\left\{\Pr\left\{\bigcup_{j}\left\{\bigcup_{l=1}^{N}\left\{\left(V_1, V_2(j)\right) \in \mathcal{A}_{l,1},\ \left(V_1, V_3(j)\right) \in \mathcal{A}_{l,2}\right\}\right\} \middle| V_1\right\}\right\}. \qquad \text{(A.4)}
\end{aligned}$$

Now, we apply the truncated union bound to the union over j, and obtain

$$\Pr\left\{\bigcup_{i,j}\left\{\bigcup_{l=1}^{N}\left\{\left(V_1\left(i\right),V_2\left(j\right)\right)\in\mathcal{A}_{l,1},\ \left(V_1\left(i\right),V_3\left(j\right)\right)\in\mathcal{A}_{l,2}\right\}\right\}\right\}$$
$$\leq L_1\mathbb{E}\left[\min\left\{1,L_2\Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1,V_2)\in\mathcal{A}_{l,1},\ (V_1,V_3)\in\mathcal{A}_{l,2}\right\}\middle|V_1\right\}\right\}\right]. \tag{A.5}$$

The third term is obtained similarly by applying the union bounds in the opposite order, and the upper bound of 1 is trivial.

The lower bound follows from de Caen's bound, which states that for for any set of events $\{\mathcal{A}_i\}_{i=1}^{M}$,

$$\Pr\left\{\bigcup_{i=1}^{M}\mathcal{A}_i\right\}\geq\sum_i\frac{\Pr\left\{\mathcal{A}_i\right\}^2}{\sum_{i'}\Pr\left\{\mathcal{A}_i\cap\mathcal{A}_{i'}\right\}}. \tag{A.6}$$

In our case, we note that by symmetry (recall that $\{V_1\left(i\right)\}_{i=1}^{L_1}$, $\{V_2\left(i\right)\}_{i=1}^{L_2}$, and $\{V_3\left(i\right)\}_{i=1}^{L_2}$ are i.i.d.), each term in the outer summation is equal, and by splitting the inner summation according to which of the (i,j) indexes coincide with (i',j'), we obtain

$$\Pr\left\{\bigcup_{i,j}\left\{\bigcup_{l=1}^{N}\left\{\left(V_1\left(i\right),V_2\left(j\right)\right)\in\mathcal{A}_{l,1},\ \left(V_1\left(i\right),V_3\left(j\right)\right)\in\mathcal{A}_{l,2}\right\}\right\}\right\}$$
$$\geq L_1L_2\Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1,V_2)\in\mathcal{A}_{l,1},\ (V_1,V_3)\in\mathcal{A}_{l,2}\right\}\right\}^2$$
$$\left((L_1-1)(L_2-1)\Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1,V_2)\in\mathcal{A}_{l,1},\ (V_1,V_3)\in\mathcal{A}_{l,2}\right\}\right\}^2\right.$$
$$+(L_2-1)\Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1,V_2)\in\mathcal{A}_{l,1},(V_1,V_3)\in\mathcal{A}_{l,2}\right\}\cap\bigcup_{l=1}^{N}\left\{\left(V_1,V_2'\right)\in\mathcal{A}_{l,1},\left(V_1,V_3'\right)\in\mathcal{A}_{l,2}\right\}\right\}$$
$$+(L_1-1)\Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1,V_2)\in\mathcal{A}_{l,1},(V_1,V_3)\in\mathcal{A}_{l,2}\right\}\cap\bigcup_{l=1}^{N}\left\{\left(V_1',V_2\right)\in\mathcal{A}_{l,1},\left(V_1',V_3\right)\in\mathcal{A}_{l,2}\right\}\right\}$$
$$\left.+\Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1,V_2)\in\mathcal{A}_{l,1},\ (V_1,V_3)\in\mathcal{A}_{l,2}\right\}\right\}\right)^{-1} \tag{A.7}$$
$$\geq L_1L_2\Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1,V_2)\in\mathcal{A}_{l,1},\ (V_1,V_3)\in\mathcal{A}_{l,2}\right\}\right\}^2$$
$$\left[4\max\left\{L_1L_2\Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1,V_2)\in\mathcal{A}_{l,1},\ (V_1,V_3)\in\mathcal{A}_{l,2}\right\}\right\}^2,\right.\right.$$
$$L_2\Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1,V_2)\in\mathcal{A}_{l,1},(V_1,V_3)\in\mathcal{A}_{l,2}\right\}\cap\bigcup_{l=1}^{N}\left\{\left(V_1,V_2'\right)\in\mathcal{A}_{l,1},\left(V_1,V_3'\right)\in\mathcal{A}_{l,2}\right\}\right\},$$

$$L_1 \Pr\left\{\bigcup_{l=1}^{N}\{(V_1,V_2)\in\mathcal{A}_{l,1},(V_1,V_3)\in\mathcal{A}_{l,2}\}\cap\bigcup_{l=1}^{N}\left\{\left(V_1',V_2\right)\in\mathcal{A}_{l,1},\left(V_1',V_3\right)\in\mathcal{A}_{l,2}\right\}\right\},$$
$$\left.\left.\Pr\left\{\bigcup_{l=1}^{N}\{(V_1,V_2)\in\mathcal{A}_{l,1},\ (V_1,V_3)\in\mathcal{A}_{l,2}\}\right\}\right\}\right]^{-1} \tag{A.8}$$

which concludes the proof. ■

Next, we prove the following result, which is a simpler version of Lemma 1.

Lemma 5 Let $\{V_1(i)\}_{i=1}^{L_1}$, $\{V_2(i)\}_{i=1}^{L_2}$, and $\{V_3(i)\}_{i=1}^{L_2}$ be independent sequences of independently and identically distributed (i.i.d.) random variables on the alphabets $\mathcal{V}_1\times\mathcal{V}_2\times\mathcal{V}_3$, respectively, with $V_1(i)\sim P_{V_1}$, $V_2(i)\sim P_{V_2}$, and $V_3(i)\sim P_{V_3}$. Fix a sequence of sets $\{\mathcal{A}_{i,1}\}_{i=1}^{N}$ and $\{\mathcal{A}_{i,2}\}_{i=1}^{N}$, where $\mathcal{A}_{i,1}\subseteq\mathcal{V}_1\times\mathcal{V}_2$ and $\mathcal{A}_{i,2}\subseteq\mathcal{V}_1\times\mathcal{V}_3$ for all $1\le i\le N$, and define

$$\mathcal{B}_{l,1}\triangleq\{v_1:\ (v_1,v_2)\in\mathcal{A}_{l,1},\ (v_1,v_3)\in\mathcal{A}_{l,2}\text{ for some }\ v_2,v_3\}, \tag{A.9}$$

and

$$\mathcal{B}_{l,2}\triangleq\{(v_2,v_3):\ (v_1,v_2)\in\mathcal{A}_{l,1},\ (v_1,v_3)\in\mathcal{A}_{l,2},\ \text{for some }\ v_1\}, \tag{A.10}$$

for $l=1,2,\ldots,N$. Then,

1) A general upper bound is given by

$$\Pr\left\{\bigcup_{i,j}\left\{\bigcup_{l=1}^{N}\{(V_1(i),V_2(j))\in\mathcal{A}_{l,1},\ (V_1(i),V_3(j))\in\mathcal{A}_{l,2}\}\right\}\right\}$$
$$\le\min\left\{1,L_1\Pr\left\{\bigcup_{l=1}^{N}\{V_1\in\mathcal{B}_{l,1}\}\right\},L_2\Pr\left\{\bigcup_{l=1}^{N}\{(V_2,V_3)\in\mathcal{B}_{l,2}\}\right\},\right.$$
$$\left.L_1L_2\Pr\left\{\bigcup_{l=1}^{N}\{(V_1,V_2)\in\mathcal{A}_{l,1},\ (V_1,V_3)\in\mathcal{A}_{l,2}\}\right\}\right\} \tag{A.11}$$

with $(V_1,V_2,V_3)\sim P_{V_1}\times P_{V_2}\times P_{V_3}$.

2) If $\{V_1(i)\}_{i=1}^{L_1}$ are pairwise independent, $\{V_2(i)\}_{i=1}^{L_2}$ are pairwise independent, $\{V_3(i)\}_{i=1}^{L_2}$ are pairwise independent, and

$$\Pr\left\{\bigcup_{l=1}^{N}\{(v_1,V_2)\in\mathcal{A}_{l,1},\ (v_1,V_3)\in\mathcal{A}_{l,2}\}\right\} \tag{A.12}$$

is the same for all $v_1\in\mathcal{B}_{1,1}$, and for all $v_1\in\mathcal{B}_{2,1}$, and so on till $v_1\in\mathcal{B}_{N,1}$, but may be different for different $\mathcal{B}_{l,1}$, and

$$\Pr\left\{\bigcup_{l=1}^{N}\{(V_1,v_2)\in\mathcal{A}_{l,1},\ (V_1,v_3)\in\mathcal{A}_{l,2}\}\right\} \tag{A.13}$$

is the same for all $(v_2, v_3) \in \mathcal{B}_{1,2}$, and so on till $(v_2, v_3) \in \mathcal{B}_{N,2}$, but may be different for different $\mathcal{B}_{l,2}$, then

$$\Pr\left\{\bigcup_{i,j}\left\{\bigcup_{l=1}^{N}\left\{\left(V_1\left(i\right), V_2\left(j\right)\right) \in \mathcal{A}_{l,1},\ \left(V_1\left(i\right), V_3\left(j\right)\right) \in \mathcal{A}_{l,2}\right\}\right\}\right\}$$
$$\geq \frac{1}{4}\min\left\{1, L_1 \Pr\left\{\bigcup_{l=1}^{N}\left\{V_1 \in \mathcal{B}_{l,1}\right\}\right\}, L_2 \Pr\left\{\bigcup_{l=1}^{N}\left\{(V_2, V_3) \in \mathcal{B}_{l,2}\right\}\right\},\right.$$
$$\left. L_1 L_2 \Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1, V_2) \in \mathcal{A}_{l,1},\ (V_1, V_3) \in \mathcal{A}_{l,2}\right\}\right\}\right\}. \tag{A.14}$$

Proof of Lemma 5: We start with the first item. To obtain (A.11) we weaken (A.1) as follows. The second term in (A.11) follows from the following fact

$$L_1 \min\left\{1, L_2 \Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1, V_2) \in \mathcal{A}_{l,1},\ (V_1, V_3) \in \mathcal{A}_{l,2}\right\}\middle| V_1\right\}\right\}$$
$$= \mathcal{I}\left\{\bigcup_{l=1}^{N}\left\{V_1 \in \mathcal{B}_{l,1}\right\}\right\} L_1 \min\left\{1, L_2 \Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1, V_2) \in \mathcal{A}_{l,1},\ (V_1, V_3) \in \mathcal{A}_{l,2}\right\}\middle| V_1\right\}\right\}$$
$$+ \mathcal{I}\left\{\bigcap_{l=1}^{N}\left\{V_1 \notin \mathcal{B}_{l,1}\right\}\right\} L_1 \min\left\{1, L_2 \Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1, V_2) \in \mathcal{A}_{l,1},\ (V_1, V_3) \in \mathcal{A}_{l,2}\right\}\middle| V_1\right\}\right\}$$
$$= \mathcal{I}\left\{\bigcup_{l=1}^{N}\left\{V_1 \in \mathcal{B}_{l,1}\right\}\right\} L_1 \min\left\{1, L_2 \Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1, V_2) \in \mathcal{A}_{l,1},\ (V_1, V_3) \in \mathcal{A}_{l,2}\right\}\middle| V_1\right\}\right\}$$
$$\leq L_1 \mathcal{I}\left\{\bigcup_{l=1}^{N}\left\{V_1 \in \mathcal{B}_{l,1}\right\}\right\} \tag{A.15}$$

where the second equality follows from the fact that the inner term in the expectation vanishes over $\bigcap_{l=1}^{N}\left\{V_1 \notin \mathcal{B}_{l,1}\right\}$, and the third inequality follows from the fact that $\min\{1, x\} \leq 1$. The third term in (A.11) follows in a similar fashion, and the forth term follows from the fact that $\min\{1, x\} \leq x$, and thus

$$L_1 \mathbb{E}\left[\min\left\{1, L_2 \Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1, V_2) \in \mathcal{A}_{l,1},\ (V_1, V_3) \in \mathcal{A}_{l,2}\right\}\middle| V_1\right\}\right\}\right]$$
$$\leq L_1 L_2 \Pr\left\{\bigcup_{l=1}^{N}\left\{(V_1, V_2) \in \mathcal{A}_{l,1},\ (V_1, V_3) \in \mathcal{A}_{l,2}\right\}\right\}. \tag{A.16}$$

This concludes the proof of the first part. The second part of Lemma 5 follows from (A.2), and the following observation. Let us consider, for example, the second term at the r.h.s. of (A.14). First, note that

$$\frac{\Pr\left\{\bigcup_{l=1}^{N}\{(V_1,V_2)\in\mathcal{A}_{l,1},\ (V_1,V_3)\in\mathcal{A}_{l,2}\}\right\}^2}{\Pr\left\{\bigcup_{l=1}^{N}\{(V_1,V_2)\in\mathcal{A}_{l,1},(V_1,V_3)\in\mathcal{A}_{l,2}\}\cap\bigcup_{l=1}^{N}\{(V_1,V_2')\in\mathcal{A}_{l,1},(V_1,V_3')\in\mathcal{A}_{l,2}\}\right\}}$$
$$=\frac{\Pr\left[\mathcal{F}\right]\Pr\left\{\bigcup_{l=1}^{N}\{(V_1,V_2)\in\mathcal{A}_{l,1},\ (V_1,V_3)\in\mathcal{A}_{l,2}\}\middle|\,\mathcal{F}\right\}^2}{\Pr\left\{\bigcup_{l=1}^{N}\{(V_1,V_2)\in\mathcal{A}_{l,1},(V_1,V_3)\in\mathcal{A}_{l,2}\}\cap\bigcup_{l=1}^{N}\{(V_1,V_2')\in\mathcal{A}_{l,1},(V_1,V_3')\in\mathcal{A}_{l,2}\}\middle|\,\mathcal{F}\right\}}. \quad \text{(A.17)}$$

where $\mathcal{F}\triangleq\bigcup_{l=1}^{N}\{V_1\in\mathcal{B}_{l,1}\}$. Now, by the additional assumptions in the second part of Lemma 5, we have

$$\Pr\left\{\bigcup_{l=1}^{N}\{(V_1,V_2)\in\mathcal{A}_{l,1},\ (V_1,V_3)\in\mathcal{A}_{l,2}\}\middle|\,\mathcal{F}\right\}$$
$$=\begin{cases}\Pr\left\{\bigcup_{l=1}^{N}\{(v_1,V_2)\in\mathcal{A}_{l,1},\ (v_1,V_3)\in\mathcal{A}_{l,2}\}\right\}, & v_1\in\mathcal{B}_{1,1}\\ \Pr\left\{\bigcup_{l=1}^{N}\{(v_1,V_2)\in\mathcal{A}_{l,1},\ (v_1,V_3)\in\mathcal{A}_{l,2}\}\right\}, & v_1\in\mathcal{B}_{2,1}\\ \quad\cdots\cdots & \\ \Pr\left\{\bigcup_{l=1}^{N}\{(v_1,V_2)\in\mathcal{A}_{l,1},\ (v_1,V_3)\in\mathcal{A}_{l,2}\}\right\}, & v_1\in\mathcal{B}_{N,1}\end{cases}. \quad \text{(A.18)}$$

Similarly,

$$\Pr\left\{\bigcup_{l=1}^{N}\{(V_1,V_2)\in\mathcal{A}_{l,1},(V_1,V_3)\in\mathcal{A}_{l,2}\}\cap\bigcup_{l=1}^{N}\left\{\left(V_1,V_2'\right)\in\mathcal{A}_{l,1},\left(V_1,V_3'\right)\in\mathcal{A}_{l,2}\right\}\middle|\,\mathcal{F}\right\}$$
$$=\begin{cases}\Pr\left\{\bigcup_{l=1}^{N}\{(v_1,V_2)\in\mathcal{A}_{l,1},\ (v_1,V_3)\in\mathcal{A}_{l,2}\}\right\}^2, & v_1\in\mathcal{B}_{1,1}\\ \Pr\left\{\bigcup_{l=1}^{N}\{(v_1,V_2)\in\mathcal{A}_{l,1},\ (v_1,V_3)\in\mathcal{A}_{l,2}\}\right\}^2, & v_1\in\mathcal{B}_{2,1}\\ \quad\cdots\cdots & \\ \Pr\left\{\bigcup_{l=1}^{N}\{(v_1,V_2)\in\mathcal{A}_{l,1},\ (v_1,V_3)\in\mathcal{A}_{l,2}\}\right\}^2, & v_1\in\mathcal{B}_{N,1}\end{cases}. \quad \text{(A.19)}$$

Thus, on substituting (A.18) and (A.19) in (A.17), we obtain

$$\frac{\Pr\left\{\bigcup_{l=1}^{N}\{(V_1,V_2)\in\mathcal{A}_{l,1},\ (V_1,V_3)\in\mathcal{A}_{l,2}\}\right\}^2}{\Pr\left\{\bigcup_{l=1}^{N}\{(V_1,V_2)\in\mathcal{A}_{l,1},(V_1,V_3)\in\mathcal{A}_{l,2}\}\cap\bigcup_{l=1}^{N}\{(V_1,V_2')\in\mathcal{A}_{l,1},(V_1,V_3')\in\mathcal{A}_{l,2}\}\right\}}$$
$$=\Pr\left[\bigcup_{l=1}^{N}\{V_1\in\mathcal{B}_{l,1}\}\right]. \quad \text{(A.20)}$$

Finally, the third term at the r.h.s. of (A.14) follows in a similar fashion. ■

Remark 2 Note that in the above results, the number of events, N, can be arbitrarily large, and in particular, exponential, without affecting the tightness of the lower and upper bounds.

Finally, note that Lemma 5 remains true for any number of sequences $\{V_1(i)\}_{i=1}^{L_1}$, $\{V_2(i)\}_{i=1}^{L_2}$,..., $\{V_K(i)\}_{i=1}^{L_2}$, and we can easily obtain a similar (exponentially tight) upper and lower bounds. Specifically, we prove the following lemma which exactly fits the structure of the probability in (66). The following result will be used in the proof of Lemma 1.

Lemma 6 Let $\{V_1(i)\}_{i=1}^{L_1}, \{V_2(i)\}_{i=1}^{L_2}, \ldots, \{V_K(i)\}_{i=1}^{L_2}$ be independent sequences of independently and identically distributed (i.i.d.) random variables on the alphabets $\mathcal{V}_1 \times \mathcal{V}_2 \times \ldots \times \mathcal{V}_K$, respectively, with $V_1(i) \sim P_{V_1}, V_2(i) \sim P_{V_2}, \ldots, V_K(i) \sim P_{V_K}$. Fix a sequence of sets $\{\mathcal{A}_{i,1}\}_{i=1}^N, \{\mathcal{A}_{i,2}\}_{i=1}^N, \ldots, \{\mathcal{A}_{i,K-1}\}_{i=1}^N$, where $\mathcal{A}_{i,j} \subseteq \mathcal{V}_1 \times \mathcal{V}_{j+1}$, for $1 \le j \le K-1$ and for all $1 \le i \le N$. Also, fix a set $\{\mathcal{A}_{i,0}\}_{i=1}^N$ where $\mathcal{A}_{i,0} \subseteq \mathcal{V}_1$ for all $1 \le i \le N$, and another sequence of sets $\{\mathcal{G}_{i,2}\}_{i=1}^N, \{\mathcal{G}_{i,3}\}_{i=1}^N, \ldots, \{\mathcal{G}_{i,K}\}_{i=1}^N$, where $\mathcal{G}_{i,j} \subseteq \mathcal{V}_j$, for $2 \le j \le K$ and for all $1 \le i \le N$. We have

$$\begin{aligned}
&\Pr\left\{\bigcup_{i,j}\left\{\bigcup_{l=1}^{N}\left\{V_1(i) \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1(i), V_{k+1}(j)) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k(j) \in \mathcal{G}_{l,k}\right\}\right\}\right\} \\
&\le \min\left\{1, L_1\mathbb{E}\left[\min\left\{1, L_2\Pr\left\{\bigcup_{l=1}^{N}\left\{V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\middle| V_1\right\}\right\}\right],\right. \\
&\left. L_2\mathbb{E}\left[\min\left\{1, L_2\Pr\left\{\bigcup_{l=1}^{N}\left\{V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\middle| \{V_k\}_{k=2}^{K}\right\}\right\}\right]\right\}
\end{aligned} \tag{A.21}$$

with $(V_1, \ldots, V_K) \sim P_{V_1} \cdots \times P_{V_K}$. Also, If $\{V_1(i)\}_{i=1}^{L_1}, \{V_2(i)\}_{i=1}^{L_2}, \ldots, \{V_K(i)\}_{i=1}^{L_2}$ are each pairwise independent, then

$$\begin{aligned}
&\Pr\left\{\bigcup_{i,j}\left\{\bigcup_{l=1}^{N}\left\{V_1(i) \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1(i), V_{k+1}(j)) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k(j) \in \mathcal{G}_{l,k}\right\}\right\}\right\} \\
&\ge \frac{1}{4}\min\left\{1, L_1\frac{\Pr\left\{\bigcup_{l=1}^{N}\left\{V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\right\}^2}{\Pr\{\mathcal{U}_1\}},\right. \\
&\quad L_2\frac{\Pr\left\{\bigcup_{l=1}^{N}\left\{V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\right\}^2}{\Pr\{\mathcal{U}_2\}}, \\
&\quad \left. L_1L_2\Pr\left\{\bigcup_{l=1}^{N}\left\{V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\right\}\right\}
\end{aligned} \tag{A.22}$$

where

$$\mathcal{U}_1 \triangleq \bigcup_{l=1}^{N} \left\{ V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1} (V_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k} \right\}$$

$$\cap \bigcup_{l=1}^{N} \left\{ V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1} \left(V_1, V'_{k+1}\right) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V'_k \in \mathcal{G}_{l,k} \right\}, \tag{A.23}$$

and

$$\mathcal{U}_2 \triangleq \bigcup_{l=1}^{N} \left\{ V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1} (V_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k} \right\}$$

$$\cap \bigcup_{l=1}^{N} \left\{ V'_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1} \left(V'_1, V_{k+1}\right) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k} \right\} \tag{A.24}$$

with $(V_1, V'_1, \ldots, V_K, V'_K) \sim P_{V_1}(v_1) \times P_{V_1}(v'_1) \times \ldots \times P_{V_K}(v_k) \times P_{V_K}(v'_k)$.

Proof of Lemma 6: The proof is exactly the same as the proof of Lemma 4. In the following, we derive, for example, the upper bound. The second term in (A.21) follows by first applying the union bound over i

$$\Pr\left\{ \bigcup_{i,j} \left\{ \bigcup_{l=1}^{N} \left\{ V_1(i) \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1} (V_1(i), V_{k+1}(j)) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k(j) \in \mathcal{G}_{l,k} \right\} \right\} \right\}$$

$$\leq L_1 \Pr\left\{ \bigcup_{j} \left\{ \bigcup_{l=1}^{N} \left\{ V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1} (V_1, V_{k+1}(j)) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k(j) \in \mathcal{G}_{l,k} \right\} \right\} \right\} \tag{A.25}$$

$$\leq L_1 \mathbb{E}\left\{ \Pr\left\{ \bigcup_{j} \left\{ \bigcup_{l=1}^{N} \left\{ V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1} (V_1, V_{k+1}(j)) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k(j) \in \mathcal{G}_{l,k} \right\} \right\} \middle| V_1 \right\} \right\}. \tag{A.26}$$

Now, we apply the truncated union bound to the union over j, and obtain

$$\Pr\left\{ \bigcup_{i,j} \left\{ \bigcup_{l=1}^{N} \left\{ V_1(i) \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1} (V_1(i), V_{k+1}(j)) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k(j) \in \mathcal{G}_{l,k} \right\} \right\} \right\}$$

$$\leq L_1 \mathbb{E}\left[\min\left\{ 1, L_2 \Pr\left\{ \bigcup_{l=1}^{N} \left\{ V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1} (V_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k} \right\} \right\} \middle| V_1 \right\} \right]. \tag{A.27}$$

The third term is obtained similarly by applying the union bounds in the opposite order, and the upper bound of 1 is trivial. The lower bound follows from de Caen's bound, as in the proof of Lemma 4 (see, (A.6)-(A.8)). ■

We are now in a position to prove Lemma 1.

Proof of Lemma 1: We start with the first item. To obtain (69) we weaken (A.21) as follows. Let $\mathcal{F} \triangleq \bigcup_{l=1}^{N} \{V_1 \in \mathcal{B}_{l,1}\}$. The second term in (69) follows from the following fact

$$\begin{aligned}
&\min\left\{1, L_2 \Pr\left\{\bigcup_{l=1}^{N}\left\{V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\middle| V_1\right\}\right\} \\
&= \mathcal{I}\{\mathcal{F}\}\min\left\{1, L_2 \Pr\left\{\bigcup_{l=1}^{N}\left\{V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\middle| V_1\right\}\right\} \\
&\quad + \mathcal{I}\{\mathcal{F}^c\}\min\left\{1, L_2 \Pr\left\{\bigcup_{l=1}^{N}\left\{V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\middle| V_1\right\}\right\} \\
&= \mathcal{I}\{\mathcal{F}\}\min\left\{1, L_2 \Pr\left\{\bigcup_{l=1}^{N}\left\{V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\middle| V_1\right\}\right\} \\
&\leq \mathcal{I}\{\mathcal{F}\}
\end{aligned} \tag{A.28}$$

where the second equality follows from the fact that the inner term in the expectation vanishes over $\bigcap_{l=1}^{N} \{V_1 \notin \mathcal{B}_{l,1}\}$, and the third inequality follows from the fact that $\min\{1, x\} \leq 1$. The third term in (A.11) follows in a similar fashion, and the forth term follows from the fact that $\min\{1, x\} \leq x$, and thus

$$\begin{aligned}
&L_1 \mathbb{E}\left[\min\left\{1, L_2 \Pr\left\{\bigcup_{l=1}^{N}\left\{V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\middle| V_1\right\}\right\}\right] \\
&\leq L_1 L_2 \Pr\left\{\bigcup_{l=1}^{N}\left\{V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\right\}.
\end{aligned} \tag{A.29}$$

This concludes the proof of the first part. The second part of Lemma 5 follows from (A.2), and the following observation. Let us consider, for example, the second term at the r.h.s. of (A.14). First, note that

$$\begin{aligned}
&\frac{\Pr\left\{\bigcup_{l=1}^{N}\left\{V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\right\}^2}{\Pr\{\mathcal{U}_1\}} \\
&= \frac{\Pr[\mathcal{F}]\Pr\left\{\bigcup_{l=1}^{N}\left\{V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\middle|\mathcal{F}\right\}^2}{\Pr\{\mathcal{U}_1|\mathcal{F}\}}.
\end{aligned} \tag{A.30}$$

where $\mathcal{U}_1$ is defined in (A.23). Now, by the additional assumptions in the second part of Lemma 5, we have

$$\Pr\left\{\bigcup_{l=1}^{N}\left\{V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\middle|\mathcal{F}\right\}$$

$$
= \begin{cases} \Pr\left\{\bigcup_{l=1}^{N}\left\{v_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(v_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\right\}, & v_1 \in \mathcal{B}_{1,1} \\ \Pr\left\{\bigcup_{l=1}^{N}\left\{v_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(v_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\right\}, & v_1 \in \mathcal{B}_{2,1} \\ \qquad \ldots\ldots & \\ \Pr\left\{\bigcup_{l=1}^{N}\left\{v_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(v_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\right\}, & v_1 \in \mathcal{B}_{N,1} \end{cases} . \tag{A.31}
$$

Similarly,

$$
\Pr\{\mathcal{U}_1|\mathcal{F}\}
$$

$$
= \begin{cases} \Pr\left\{\bigcup_{l=1}^{N}\left\{v_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(v_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\right\}^2, & v_1 \in \mathcal{B}_{1,1} \\ \Pr\left\{\bigcup_{l=1}^{N}\left\{v_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(v_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\right\}^2, & v_1 \in \mathcal{B}_{2,1} \\ \qquad \ldots\ldots & \\ \Pr\left\{\bigcup_{l=1}^{N}\left\{v_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(v_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\right\}^2, & v_1 \in \mathcal{B}_{N,1} \end{cases} . \tag{A.32}
$$

Thus, on substituting (A.31) and (A.32) in (A.30), we obtain

$$
\frac{\Pr\left\{\bigcup_{l=1}^{N}\left\{V_1 \in \mathcal{A}_{l,0},\ \bigcap_{k=1}^{K-1}(V_1, V_{k+1}) \in \mathcal{A}_{l,k},\ \bigcap_{k=2}^{K} V_k \in \mathcal{G}_{l,k}\right\}\right\}^2}{\Pr\{\mathcal{U}_1\}}
$$

$$
= \Pr\left[\bigcup_{l=1}^{N}\{V_1 \in \mathcal{B}_{l,1}\}\right]. \tag{A.33}
$$

Finally, the third term at the r.h.s. of (A.14) follows in a similar fashion. ■

REFERENCES

[1] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[2] E. C. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability*, vol. 1. Berkeley, CA: Univ. California Press, 1961, pp. 611–64.

[3] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *Annals Probabil.*, vol. 2, no. 5, pp. 805–814, 1974.

[4] A. B. Carleial, "Interference channels," *IEEE Trans. on Inf. Theory*, vol. IT-24, pp. 60–70, Jan. 1978.

[5] H. Sato, "Two-user communication channels," *IEEE Trans. on Inf. Theory*, vol. IT-23, pp. 295–304, May. 1977.

[6] A. B. Carleial, "A case where intereference does not reduce capacity," *IEEE Trans. on Inf. Theory*, vol. IT-21, pp. 569–570, Sep. 1975.

[7] R. Benzel, "The capacity region of a class of discrete additive degraded interference channels," *IEEE Trans. on Inf. Theory*, vol. IT-25, pp. 228–231, Mar. 1979.

[8] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. on Inf. Theory*, vol. IT, no. 27, pp. 49–60, Jan. 1981.

[9] C. Nair, L. Xia, and M. Yazdanpanah, "Sub-optimality of the Han-Kobayashi achievable region for interference channels," *submitted to ISIT 2015*, Feb. 2015. [Online]. Available: http://arxiv.org/abs/1502.02589

[10] R. Etkin, N. Merhav, and E. Ordentlich, "Error exponents of optimum decoding for the interference channel," *IEEE Trans. on Inf. Theory*, vol. 56, no. 1, pp. 40–56, Jan. 2010.

[11] C. Chang, R. Etkin, and E. Ordentlich, "Interference channel capacity region for randomized fixed-composition codes," *HP Labs Technical Report*, 2009. [Online]. Available: http://www.hpl.hp.com/techreports/2008/HPL-2008-194R1.html

[12] J. Scarlett, A. Martinez, and A. G. Fábregas, "Multiuser coding techniques for mismatched decoding," *submitted to IEEE Trans. on Inf. Theory*, Nov. 2013. [Online]. Available: arxiv.org/pdf/1311.6635

[13] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[14] I. Csiszár and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 2011.

[15] N. Merhav, "Relations between random coding exponents and the statistical physics of random codes," *IEEE Trans. on Inf. Theory*, vol. 55, no. 1, pp. 83–92, Jan. 2009.

[16] ——, "Exact random coding exponents of optimal bin index decoding," *IEEE Trans. on Inf. Theory*, vol. 60, no. 10, pp. 6024–6031, Oct. 2014.

[17] ——, "List decoding-random coding exponents and expurgated exponents," *IEEE Trans. on Inf. Theory*, vol. 60, no. 11, pp. 6749–6759, Nov. 2014.

[18] ——, "Exact correct-decoding exponent for the wiretap channel decoder," *IEEE Trans. on Inf. Theory*, vol. 60, no. 12, pp. 7606–7615, Dec. 2014.

[19] W. Huleihel, N. Weinberger, and N. Merhav, "Erasure/list random coding error exponents are not universally achievable," *submitted to IEEE Trans. on Inf. Theory*, Oct. 2014. [Online]. Available: http://arxiv.org/abs/1410.7005

[20] N. Merhav, "The generalized random energy model and its application to the statistical physics of ensembles of hierarchical codes," *IEEE Trans. on Inf. Theory*, vol. 55, no. 3, pp. 1250–1268, May 2009.

[21] N. Shulman, "*Communication over an unknown channel via common broadcasting*," Ph.D. dissertation, Tel-Aviv University, 2003, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.407.7542&rep=rep1&type=pdf.

[22] B. Bandemer, A. El Gamal, and Y. K. Kim, "Optimal achievable rates for interference networks with random codes," *submitted to IEEE Trans. on Inf. Theory*, 2012. [Online]. Available: http://arxiv.org/abs/1210.4596

[23] L. Weng, S. S. Pradhan, and A. Anastasopoulos, "Error exponent regions for Gaussian broadcast and multiple-access channels," *IEEE Trans. on Inf. Theory*, vol. 54, no. 7, pp. 2919–2942, July 2008.